CORRECTIONS TO THE KDV APPROXIMATION FOR WATER WAVES

J. DOUGLAS WRIGHT

ABSTRACT. In order to investigate corrections to the common KdV approximation for surface water waves in a canal, we derive modulation equations for the evolution of long wavelength initial data. We work in Lagrangian coordinates. The equations which govern corrections to the KdV approximation consist of linearized and inhomogeneous KdV equations plus an inhomogeneous wave equation. These equations are explicitly solvable and we prove estimates showing that they do indeed give a significantly better approximation than the KdV equation alone.

AMS classification: 76B15, 35Q51, 35Q53

1. INTRODUCTION

It is often easier to write down a partial differential equation which models a physical phenomena than it is to study solutions of such an equation. Equations which model the evolution of the surface of a fluid in a canal have been known since at least the 19th century, however it has only been in recent years that questions of existence and uniqueness for general initial data have been answered (see [38], [39] and [23]). Moreover, numerical simulations of water waves and similarly complex phenomena are frequently time consuming and challenging to implement. Consequently, it can be quite difficult to say much about the behavior of a general solution. And so scientists often restrict their attention to limiting cases—for instance, one may assume that solutions are of long wavelength and small amplitude (see Figure 1). Under such a supposition, a *modulation equation* may be (formally) derived. In particular, one hopes that the modulation equation:

- is well-posed,
- is either explicitly solvable or easy to solve numerically and
- captures the essential behavior of the original system.

Remarkably, many seemingly disparate physical phenomena possess modulation equations of the same form. For solutions of long wavelength, Korteweg-de Vries (KdV) equations are often used as modulation equations for a wide variety of non-linear dispersive systems, including the water wave equation, the Euler-Poisson equations for plasma dynamics and the Fermi-Pasta-Ulam equation for the interaction of particles in an infinite lattice.

Despite the fact that modulation equations have been in use for over a hundred years—the KdV equation was first proposed as a model for water
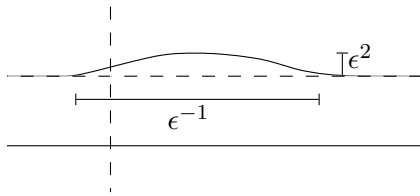
FIGURE 1. The long wave, small amplitude scaling.

waves by Boussinesq in 1872 and also by Korteweg and de Vries in 1895—only recently have attempts been made to rigorously connect the behavior of the modulation equations to the original physical problem. In particular, through the work of Craig [9], Kano and Nishida [17] [18], Kalyakin [16], Schneider [30], Ben Youssef and Colin [1] and Schneider and Wayne [31], [33], the validity of KdV equations as a leading order approximation to the evolution of long wavelength water waves and to a number of other dispersive partial differential equations has been established.

In many respects, the KdV equation is an ideal modulation equation; it is simple in form and explicitly solvable *via* the inverse scattering transform. Nevertheless, both experimentally and numerically one observes deviations from the predictions of the KdV approximation. In this paper we derive a hierarchy of modulation equations which govern corrections to the KdV model and also prove rigorously that these higher order equations do indeed improve the accuracy of the approximation. While the correction is valid in general long wavelength/small amplitude settings, heuristically the model is set up to better approximate interactions between solitary waves—both counterpropagating collisions and unidirectional interactions.

Note that in [36], as a case study, Wayne and Wright examined higher order corrections to the KdV approximation to a Boussinesq equation. As the KdV equation is in some sense a universal approximation for long waves, we expect that the equations for corrections to this approximation will also be universal. Indeed, our results show that the higher order corrections for the water wave equation are nearly identical to those for the Boussinesq equation. Also, since a significant part of the work for the Boussinesq problem consists of showing that the modulation equations have well-behaved solutions over the time scales of interest, this is of use in tackling the water wave problem.

We now describe our results in some detail. The equations of motion for a water wave in an infinitely long canal (commonly called *the water wave equation*) are

$$x_{tt}(1 + x_\alpha) + y_\alpha(1 + y_{tt}) = 0$$

(WW)
$$y_t = K(x, y)x_t$$

$$\alpha \in \mathbb{R},\ t \geq 0,\ (x(\alpha, t), y(\alpha, t)) \in \mathbb{R}^2,$$
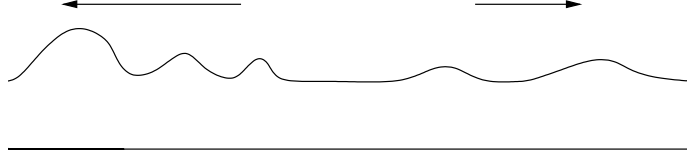
FIGURE 2. Sketch of the right and left moving wave trains.

where $K(x, y)$ is a complicated operator (see Section 3 for the definition of K) and $(\alpha + x(\alpha, t), y(\alpha, t))$ parameterizes the free surface. According to the KdV approximation results of [31], to the order of the approximation long wavelength solutions of (WW) split up into two pieces, one a right moving wave train and one a left moving wave train. Each of these wave trains evolves according to a KdV equation, and *there is no interaction between the left and right moving pieces.* That is, for $0 < \epsilon \ll 1$, if we scale amplitudes to be $O(\epsilon^2)$ (*i.e.* small) and wavelengths to be $O(\epsilon^{-1})$ (*i.e.* long), then for times of $O(\epsilon^{-3})$, solutions to (WW) satisfy

$$(1) \qquad -x_\alpha(\alpha, t) = \epsilon^2 U(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^2 V(\epsilon(\alpha + t), \epsilon^3 t) + O(\epsilon^4)$$

where U and V satisfy the KdV equations

$$(\text{KdV}) \qquad \begin{aligned} -2\partial_T U &= \frac{1}{3}\partial_{\beta_-}^3 U + \frac{3}{2}\partial_{\beta_-}(U^2) \\ 2\partial_T V &= \frac{1}{3}\partial_{\beta_+}^3 V + \frac{3}{2}\partial_{\beta_+}(V^2). \end{aligned}$$

See Figure 2. Here, $\beta_\pm = \epsilon(\beta \pm t)$ represent long wavelength moving reference frames and $T = \epsilon^3 t$ is the very long time scale coordinate. For technical reasons (which we discuss later), $-x_\alpha$ is the natural variable to estimate for the water wave equation. To lowest order, $-x_\alpha$ is proportional to the height of the wave. At higher order, this ceases to be true, though for purposes of intuition one can think of $-x_\alpha$ as representing the wave amplitude.

The KdV equation was initially derived from the water wave equation in an attempt to prove the existence of a solitary wave solution for waves in a canal. Famously, the KdV equation admits solitary wave solutions and also multi-soliton solutions. (See Figure 3). We will frequently refer to multi-soliton solutions as "overtaking wave" collisions. We remind the reader that the only notable first order effect after such a collision is that the waves are phase shifted after the collision.

Given the results in [31], one expects to see similar behavior in solutions systems modeled by KdV equations. Though it is unknown if these soliton-like solutions persist globally, analogous behavior is indeed observed for very long times (see [14]). The most notable deviation between true solutions and the KdV approximation is the size of the phase shift after a collision. In addition, soliton-like solutions to the type of systems we study frequently develop a very small amplitude dispersive wave train behind each soliton,
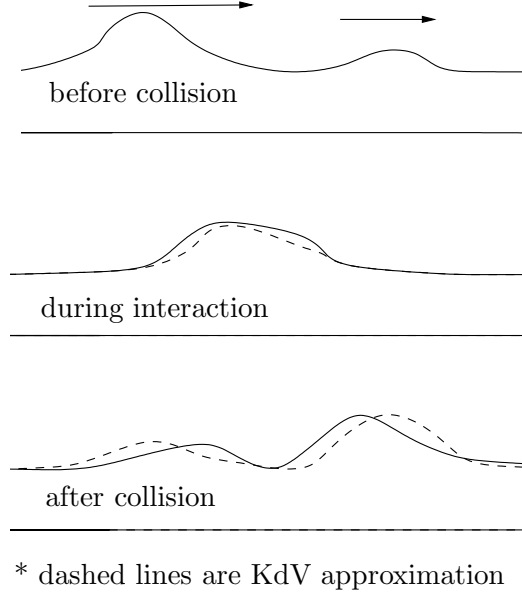
before collision

during interaction

after collision

* dashed lines are KdV approximation

FIGURE 3. Sketch of the overtaking wave interaction.
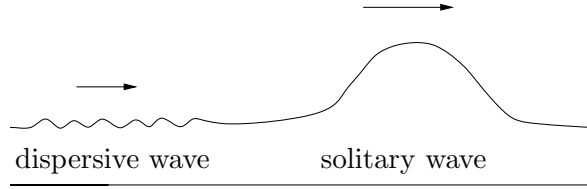

dispersive wave   solitary wave

FIGURE 4. Sketch of the dispersive wave.

which moves in the same direction, see Figure 4. The KdV approximation does not predict the existence of these dispersive wave trains. As these sorts of discrepancies are observed even in the case where there is only one wave train moving unidirectionally, we believe that they are, loosely, independent of interactions between the left and right moving wave trains. They reflect intrinsic differences between the approximation and the original system.

On the other hand, there is evidence that a noticeable interaction takes place between the left and right moving waves. One can see from the form of the approximation in equation (1) that during a head-on collision of waves moving in opposite directions, the KdV approximation predicts that the heights of the waves add linearly. In true head-on collisions in solutions to the water wave equation, however, the height of the waves is slightly different from the sum of the heights of the waves taken separately—it is slightly
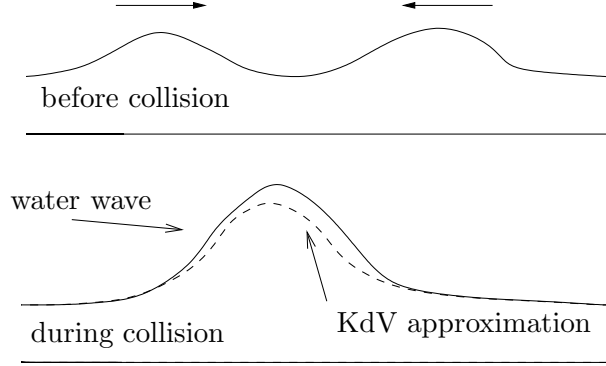
FIGURE 5. Sketch of the head-on collision.

larger. (See the works of Maxworthy [24], Byatt-Smith [6] [7], Cooker, Weidman and Bale [8], and Su and Mirie [34] [35].) We sketch this in Figure 5.

Thus we might expect two types of corrections to the KdV approximation:

- corrections due to the fact that, even in the case of a purely right (or left) moving wave train, solutions to the water wave equation are not exactly described by solutions to the KdV equation. We will refer to this source of error as *unidirectional error*.
- corrections due the fact that the left and right moving wave trains will interact at higher order. We call such errors *counterpropagation error*.

Both of these types of corrections are apparent in our results, and to incorporate these two types of corrections, we add an additional three functions to the KdV wave trains. The first two, F and G, will correct for unidirectional errors. The third, P, will correct for counterpropagation errors. We scale the amplitudes of these three functions so they are $O(\epsilon^4)$, which is the same as the order of the error in using only the KdV equations. F and G will take the same functional form as U and V, as they correct for differences between the approximate and actual wave trains. So we add

$$\epsilon^4 F(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^4 G(\epsilon(\alpha + t), \epsilon^3 t)$$

to the first order approximation (1).

We do not expect P to be moving strictly left or right, as it corrects for the interaction between waves moving in opposite directions. So its spatial dependence will be on $\beta = \epsilon\alpha$. Suppose that the functions U and V are solitary wave solutions. In the long wavelength variables we are considering, this means that the right and left moving wave packets are large only over a length of $O(\epsilon^{-1})$. In addition, the reference frame moves with unit velocity. Thus we expect any interaction of the two waves to last a time of $O(\epsilon^{-1})$. Accordingly, we let P depend on the time variable $\tau = \epsilon t$. That is we add

a correction term of the form

$$\epsilon^4 P(\epsilon\alpha, \epsilon t)$$

to the KdV model.

Through formal means we find that P satisfies an inhomogeneous wave equation

(IW) $$\partial_\tau^2 P - \partial_\beta^2 P = 3\partial_\beta^2 \left(U(\beta - \tau, \epsilon^2\tau)V(\beta + \tau, \epsilon^2\tau) \right).$$

Similarly, F and G satisfy a pair of driven, linearized KdV equations

(LK)
$$-2\partial_T F = \frac{1}{3}\partial_{\beta_-}^3 F + 3\partial_{\beta_-}(UF) + J^-$$
$$2\partial_T G = \frac{1}{3}\partial_{\beta_+}^3 G + 3\partial_{\beta_+}(VG) + J^+.$$

Notice that these equations are linearized about the KdV solutions U and V. The inhomogeneous terms J^- and J^+ are made up of a combination of sums and products of U, V and P. For explicit forms of these driving terms see the equations (22). Linearized KdV equations are explicitly solvable, though this is a complicated matter (see [27] and [15]). However, solutions are simple to compute numerically.

One can solve inhomogeneous wave equations explicitly and easily *via* the method of characteristics. Moreover, we can reduce such systems to a pair of transport equations by the following fact:

Fact 1. *If $\partial_t f - \partial_x f = 1/2\partial_x h$ and $\partial_t g + \partial_x g = -1/2\partial_x h$, then $q = f + g$ satisfies $\partial_t^2 q - \partial_x^2 q = \partial_x^2 h$.*

Thus, we have

(T)
$$P_\tau^- + P_\beta^- = -\frac{3}{2}\partial_\beta(U(\beta - \tau, \epsilon^2\tau)V(\beta + \tau, \epsilon^2\tau))$$
$$P_\tau^+ - P_\beta^+ = \frac{3}{2}\partial_\beta(U(\beta - \tau, \epsilon^2\tau)V(\beta + \tau, \epsilon^2\tau))$$

where

$$P(\beta, \tau) = P^+(\beta, \tau) + P^-(\beta, \tau).$$

We remark that the initial data for the modulation equations is determined from initial conditions for the original system in ways described in Section 6. Also, this hierarchy of higher order modulation equations is nearly identical to that derived in Wayne and Wright [36] for the Boussinesq equation—the chief difference lying in the specific forms of the inhomogeneous terms $J^\pm$.

To enforce the notion of spatial localization, we will be considering initial data which is of rapid decay, that is, initial data in

$$H^s(m) = \left\{ f(\alpha) | (1 + \alpha^2)^{m/2} f(\alpha) \in H^s \right\}.$$

The inner product on $H^s(m)$ is given by

$$(f(\cdot), g(\cdot))_{H^s(m)} = \left((1 + \cdot^2)^{m/2} f(\cdot), (1 + \cdot^2)^{m/2} g(\cdot)\right)_{H^s},$$

where we use the standard inner product in H^s. In particular, the known soliton solutions of the KdV equations are in such spaces.

That the KdV equations have solutions for all times with this sort of initial data is well known. In particular, we have from [31]

Theorem 1. *Let $\sigma \geq 4$. Then for all $C_I, T_0 > 0$ there exists $C_1 > 0$ such that if U, V satisfy (KdV) with initial conditions U_0, V_0 and*

(2) $$\max\{\|U_0\|_{H^\sigma(4) \cap H^{\sigma+4}(2) \cap H^{\sigma+9}}, \|V_0\|_{H^\sigma(4) \cap H^{\sigma+4}(2) \cap H^{\sigma+9}}\} < C_I$$

then

(3)
$$\sup_{T \in [0,T_0]} \left\{\|U(\cdot, T)\|_{H^\sigma(4) \cap H^{\sigma+4}(2) \cap H^{\sigma+8}}, \|V(\cdot, T)\|_{H^\sigma(4) \cap H^{\sigma+4}(2) \cap H^{\sigma+8}}\right\} < C_1.$$

On the other hand it is less clear that solutions of (IW) and (LK) will remain bounded over the very long time scales necessary for the KdV approximation. In [36] we proved the following result which guarantees that the solutions of the modulation equations remain bounded for sufficiently long times.

Proposition 1. *Fix $T_0 > 0$ and $\sigma > 11/2$. Suppose, U_0, V_0 satisfy (2) and U, V, $P^{\pm}$, F and G satisfy (KdV), (T) and (LK), then there exists a constant C_2, independent of ϵ, such that the solutions of (T) and (LK) satisfy the estimates below:*

$$\sup_{\tau \in [0, T_0 \epsilon^{-2}]} \|P^{\pm}(\cdot, \tau)\|_{H^{\sigma+3}} \leq C_2$$

$$\sup_{T \in [0, T_0]} \left\{\|F(\cdot, T)\|_{H^\sigma \cap H^{\sigma-4}(2)}, \|G(\cdot, T)\|_{H^\sigma \cap H^{\sigma-4}(2)}\right\} \leq C_2.$$

Moreover, $P^{\pm}(\beta, \tau) = \varphi^{\pm}(\beta_{\pm}, T)$ with

$$\sup_{T \in [0, T_0]} \|\varphi^{\pm}(\cdot, T)\|_{H^{\sigma+3} \cap H^{\sigma-1}(2)} \leq C_2.$$

Finally we note that since $\partial_\tau P = \partial_\beta P^+ - \partial_\beta P^-$ we have $\|\partial_\tau P\|_s \leq \|\partial_\beta P\|_s$.

With this preliminary result in hand we can state our principal results. Denote the sum of the modulation functions, properly scaled, as

(4)
$$\begin{aligned} -\psi^d(\alpha, t) = {} & \epsilon^2 U\left(\epsilon(\alpha - t), \epsilon^3 t\right) + \epsilon^2 V\left(\epsilon(\alpha + t), \epsilon^3 t\right) \\ & + \epsilon^4 F\left(\epsilon(\alpha - t), \epsilon^3 t\right) + \epsilon^4 G\left(\epsilon(\alpha + t), \epsilon^3 t\right) \\ & + \epsilon^4 P(\epsilon\alpha, \epsilon t). \end{aligned}$$

As mentioned earlier, (x, y) are not the natural coordinates to study solutions to (WW). The coordinates we use are $(x_\alpha, y, x_t)^{tr}$. x_α is approximated

by ψ^d, and the functions y and x_t are approximated by functions we denote ψ^y and ψ^u, respectively. They are given by

$$
\begin{aligned}
\psi^y(\alpha, t) = \ & \epsilon^2 U(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^2 V(\epsilon(\alpha + t), \epsilon^3 t) \\
& + \epsilon^4 F(\epsilon(\alpha - t), \epsilon^3 t) + \epsilon^4 G(\epsilon(\alpha + t), \epsilon^3 t) + \epsilon^4 P(\epsilon\alpha, \epsilon t) \\
& + \frac{1}{3}\epsilon^4 \partial_{\beta_-}^2 U(\epsilon(\alpha - t), \epsilon^3 t) + \frac{1}{3}\epsilon^4 \partial_{\beta_+}^2 V(\epsilon(\alpha + t), \epsilon^3 t) \\
& + \epsilon^4 \left(U(\epsilon(\alpha - t), \epsilon^3 t) + V(\epsilon(\alpha + t), \epsilon^3 t) \right)^2
\end{aligned}
$$

(5)

and

$$
\begin{aligned}
\psi^u(\alpha, t) = \ & \epsilon^2 U(\epsilon(\alpha - t), \epsilon^3 t) - \epsilon^2 V(\epsilon(\alpha + t), \epsilon^3 t) \\
& + \epsilon^4 F(\epsilon(\alpha - t), \epsilon^3 t) - \epsilon^4 G(\epsilon(\alpha + t), \epsilon^3 t) \\
& + \epsilon^4 \varphi^-(\epsilon(\alpha - t), \epsilon^3 t) - \epsilon^4 \varphi^+(\epsilon(\alpha + t), \epsilon^3 t) \\
& + \frac{1}{6}\epsilon^4 \partial_{\beta_-}^2 U(\epsilon(\alpha - t), \epsilon^3 t) - \frac{1}{6}\epsilon^4 \partial_{\beta_+}^2 V(\epsilon(\alpha + t), \epsilon^3 t) \\
& + \frac{3}{4}\epsilon^4 U^2(\epsilon(\alpha - t), \epsilon^3 t) - \frac{3}{4}\epsilon^4 V^2(\epsilon(\alpha + t), \epsilon^3 t).
\end{aligned}
$$

(6)

We discuss the origin of these equations in Section 4.

The approximation will be valid in the space

$$
\mathfrak{H}^s = H^s \times H^s \times H^{s-1/2}.
$$

Our main result is:

Theorem 2. *Fix T_0, $C_I > 0$, $s > 4$, $\sigma \geq s+7$. Suppose U, V, P, F and G satisfy equations (KdV), (IW) and (LK) and ψ^d, ψ^y and ψ^u are the combinations of these functions given in (4), (5) and (6). Then there exist $\epsilon_0 > 0$ and $C_F > 0$ such that the following is true. If the initial conditions for (WW) are of the form $((x_\alpha(\alpha, 0), y(\alpha, 0), x_t(\alpha, 0))^{tr} = \left(0, \epsilon^2 \Theta_y(\epsilon\alpha), \epsilon^2 \Theta_u(\epsilon\alpha)\right)^{tr}$ with*

$$
\max_{i=y,z} \left\{ \|\Theta_i(\cdot)\|_{H^\sigma(4) \cap H^{\sigma+4}(2) \cap H^{\sigma+9}} \right\} \leq C_I
$$

then for $\epsilon \in (0, \epsilon_0)$, there is a reparameterization of the free surface such that the unique solution to (WW) satisfies

$$
\left\| \begin{pmatrix} x_\alpha(\cdot, t) \\ y(\cdot, t) \\ x_t(\cdot, t) \end{pmatrix} - \begin{pmatrix} \psi^d(\cdot, t) \\ \psi^y(\cdot, t) \\ \psi^u(\cdot, t) \end{pmatrix} \right\|_{\mathfrak{H}^s} \leq C_F \epsilon^{11/2}
$$

for $t \in [0, T_0\epsilon^{-3}]$. The constant C_F does not depend on ϵ.

Remark 1. *The loss of the one half power of ϵ in Theorem 2 is caused by the long wave scaling and not a lack of sharpness in the estimates.*

Remark 2. *It is clear that the form of the initial conditions specified in the hypotheses of this theorem do not agree with those found by setting $t = 0$ in the approximation inequality (unless, of course, $\psi^d \equiv 0$). This is precisely why we mention the need to reparameterize the free surface. We discuss this at length in Section 6.*

Less technically, this theorem states that solutions to (WW), in the long wavelength limit, satisfy

$$x_\alpha(\alpha, t) = \psi^d(\alpha, t) + O(\epsilon^6)$$

for times of $O(\epsilon^{-3})$. This is, as expected, a marked improvement over the use of KdV alone.

We note that this is not the first time that linearized KdV equations have been put forward as a means to improve the accuracy of the KdV approximation. Other instances where linearized KdV equations appear include Sachs [28], Sattinger, Haragus and Nicholls [14], Kodama and Taniuti [21] and Drazin [11]. Moreover, there have been numerous models put forward over the years which model water waves in the same scaling regime we are considering. We refer the reader to Kodama [20], Olver [26], Bona, Pritchard and Scott [2], Craig and Groves [10], Dullin, Gottwald and Holm [12], and Bona and Chen [3]. Much of the work done in the above papers pertains to analyzing the behavior of the model equations and not to their connection to the original system. A notable exception is the recent work by Bona, Lannes and Colin [4] wherein they prove the rigorous validity of a large number of Boussinesq style models. Our particular combination of linearized KdV equations with an inhomogeneous wave equation appears to be unique and is asymptotically the most accurate model for long wavelength solutions to the water wave equation which has currently been justified rigorously.

The remainder of this paper is organized as follows. First, in Section 2, we conduct a prelimary discussion of the water wave equation. Sections 3 and 5 contain a thorough discussion of the operator $K(x, y)$. Then, in Section 4 we derive the higher order modulation equations and prove an important estimate. In Sections 6 we prove the validity of the approximation *i.e.* Theorem 2. Finally, Section 7 contains the details for a number of proofs.

2. Preliminaries

We begin by discussing the water wave problem in greater detail. Consider an infinitely long canal of unit mean depth in two-dimensions (see Figure 6). We denote the region occupied by the fluid at time t as $\Omega(t)$, and the upper surface as $\Gamma(t)$. We parameterize $\Gamma(t)$ by $(\tilde{x}(\alpha, t), y(\alpha, t))$, where $\alpha \in \mathbb{R}$ is the parameter, and $\tilde{x}$ and y are the real-valued coordinate functions. It is useful to break $\tilde{x}$ up as follows,

$$\tilde{x}(\alpha, t) = \alpha + x(\alpha, t).$$

$$\Gamma(t) = \{(\alpha + x(\alpha, t), y(\alpha, t)) \,|\, \alpha \in \mathbb{R}\}$$
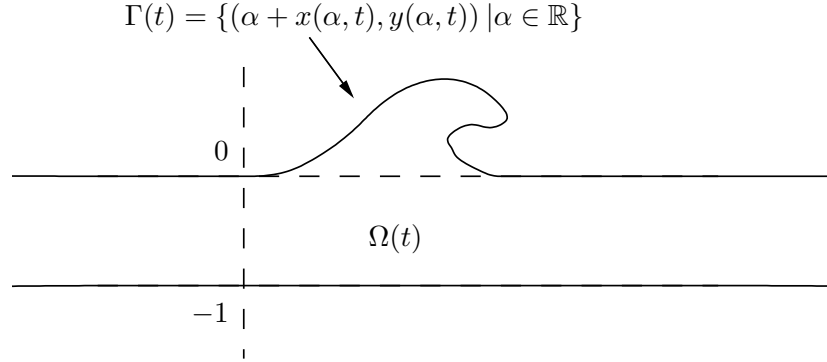

FIGURE 6. The water wave in Lagrangian coordinates.

We consider fluids which are inviscid and incompressible and flows which are irrotational. Also, we assume that the pressure on the top surface is a constant, and that the acceleration due to gravity is 1. With these assumptions, the evolution of x and y are given by the equations

(WW)
$$x_{tt}(1 + x_\alpha) + y_\alpha(1 + y_{tt}) = 0$$
$$y_t = K(x, y)x_t.$$

(See [9].)

The first of these two equations is found from Euler's equations for fluid motion. The operator K in the second line is a transformation which is linear in x_t, but depends nonlinearly on (x, y). That such an operator exists and gives a relationship between x_t and y_t is discussed in Sections 3 and 5, along with an analysis of K. Much of the difficulty in answering questions about the water wave equation is related to this operator. If the surface of the water is perfectly flat, *i.e.* $x = y = 0$, then we have $K(0, 0) = K_0$ where K_0 is a linear operator defined by $\widehat{K_0 f}(k) = \widehat{K_0}(k)\widehat{f}(k)$, with $\widehat{K_0}(k) = -i\tanh(k)$. Notice that since tanh is a bounded function, K_0 is bounded from H^s to H^s. K_0 will be appearing frequently.

This formulation of the water wave problem is said to be in *Lagrangian*, or *material*, coordinates. In this point of view we are not in a fixed "lab" frame, but instead we are tracking the position of each "particle" of water separately. That is, $(\alpha + x(\alpha, t), y(\alpha, t))$ gives the location of the particle which was initially at $(\alpha + x(\alpha, 0), y(\alpha, 0))$. The *laboratory*, or *Eulerian*, point of view is to fix a system of coordinates on the fluid domain and to measure the velocity of the fluid at each point of this fixed reference frame. For our purposes, it is far more convenient to use Lagrangian coordinates, however experimentalists work with Eulerian coordinates. We will give formulae for the approximation in terms of Eulerian coordinates in a future publication. The interested reader may also see the author's thesis [37] for this information.

Since the water wave equation is second order in time for both x and y, one might suppose that four functions are neccessary to specify the initial state of the system—$x(\alpha, 0)$, $y(\alpha, 0)$, $x_t(\alpha, 0)$ and $y_t(\alpha, 0)$. In fact, in general only three are needed, and for initial data with small amplitudes, only two. The relationship $y_t = K(x, y)x_t$ specifies the value of $y_t(\alpha, 0)$ given the other three functions. We can also "do away" with the initial condition for x, provided we are in the small amplitude, long wave limit. If $x(\alpha, 0)$ and its first derivative are sufficiently small, then $\alpha + x(\alpha, 0)$ will be invertible. This implies that $\Gamma(0)$ will be a graph over the horizontal coordinate. And so, without loss of generality, we can reparameterize the initial conditions so that

$$\Gamma(0) = \{(\alpha, y(\alpha, 0)) | \alpha \in \mathbb{R}\}.$$

Thus we need only to choose $y(\alpha, 0)$ and $x_t(\alpha, 0)$. As it turns out, we need to reparameterize the system one more time to prove the approximation theorem, but we will leave this technicality until Section 6. The essential point here is that due to the freedom in choosing the initial parameterization, we can eliminate two of our initial conditions.

Even though we can assume that $x(\alpha, 0) = 0$ (or, alternately, is small), this coordinate grows linearly in time. (See the linear estimates in Chapter 2 of [31].) As we are concerned with very long time scales, this is a problem. As was shown in [31], one can replace x with the new coordinate $z = K_0 x$, which is well-behaved over long times. Rewriting (WW) with this new variable, we have

$$\text{(WW3)} \qquad \begin{aligned} \partial_t z &= K_0 u \\ \partial_t y &= K(z, y)u \\ \partial_t u &= -\frac{\partial_\alpha y(1 + \partial_t^2 y)}{1 + K_0^{-1}\partial_\alpha z}. \end{aligned}$$

We will see that the operator $K(x, y)$ in truth depends not on x but on z, so the abuse of the notation $K(z, y)$ above is in some sense legitimate (see Section 3 for further discussion). Furthermore, even though K_0^{-1} is not well-defined, as it blows up at frequency $k = 0$, the composition

$$L = K_0^{-1}\partial_\alpha$$

is well-defined as its symbol, $-k/\tanh(k)$, has no singularities; L is also invertible. Finally, we notice that the Maclaurin expansion of $\widehat{K_0}(k) = -ik + O(k^3)$. Thus to lowest order $K_0 \sim -\partial_\alpha$. And so $z \sim -x_\alpha$. This is precisely the reason why, in the Introduction, we stated that $-x_\alpha$ is a natural coordinate for the water wave equation.

Though we will primarily be working with the three-dimensional system WW3, we will need to embed this system into a four-dimensional system to prove certain aspects of the validity of the approximation. We introduce

the new coordinate $a = u_t$, and (WW3) becomes

$$\partial_t z = K_0 u$$
$$\partial_t y = K(z, y) u$$
(WW4)
$$\partial_t u = a$$
$$\partial_t a = -\frac{a \partial_\alpha u + \partial_\alpha \partial_t (1 + \partial_t^2 y) + \partial_t y \partial_t^3 y}{1 + K_0^{-1} \partial_\alpha z}$$

Though things appear to be getting out of hand, we remark that this is as large a system as we will need. Results in [31] prove that solutions to equations (WW), (WW3) and (WW4) do indeed exist for long times. We will be considering solutions to the four dimensional system which are in

$$\mathfrak{H}_e^s = H^s \times H^s \times H^{s-1/2} \times H^{s-1}.$$

The main goal of this paper is to prove Theorem 2. To do this, we first prove a similar theorem for solutions to (WW3), from which Theorem 2 will follow. Let

(7)
$$\Psi^d(\alpha, t) = \psi^d(\alpha, t) + \epsilon^6 W_3(\epsilon \alpha, \epsilon t).$$

The additional function W_3 will solve an equation we specify later. While we assure the reader that we will be following Chekhov's rule and that this gun which appears in the first act will be fired in the third, interested parties may look ahead to equation (23) in Section 4 for more information about W_3. Define the functions

$$\Psi^z = L^{-1} \Psi^d$$
$$\Psi^y = \Psi^z + \epsilon^4 \Delta_1 + \epsilon^6 \Delta_2$$
$$\Psi^u = \partial_\alpha^{-1} \partial_t \Psi^d.$$

Δ_1 and Δ_2 are combinations of solutions to the modulation equations and are given in equations (17) and (19) in Section 4. We justify the presense of the inverse derivative in Ψ^u by means of example. Ψ^d contains the term U which solves a KdV equation. Thus $\partial_t \Psi^d$ contains the terms from the right hand side of the KdV equation, all of which are perfect space derivatives. These then are cancelled by the inverse derivative. It is simple to check that this method applies to all terms in Ψ^d.

With these function, we have

Theorem 3. *Fix T_0, $C_I > 0$, $s > 4$, $\sigma \geq s + 7$. Let Ψ^d, Ψ^z, Ψ^y and Ψ^u be as above. Moreover assume the initial conditions for (KdV) satisfy*

$$\max\{\|U_0\|_{H^\sigma(4) \cap H^{\sigma+4}(2) \cap H^{\sigma+9}}, \|V_0\|_{H^\sigma(4) \cap H^{\sigma+4}(2) \cap H^{\sigma+9}}\} < C_I.$$

Then there exist $\epsilon_0 > 0$ and $C_F = C_F(T_0, C_I, s) > 0$ such that if the initial conditions for (WW3) are of the form

$$\begin{pmatrix} z(\alpha, 0) \\ y(\alpha, 0) \\ u(\alpha, 0) \end{pmatrix} = \begin{pmatrix} \Psi^z(\alpha, 0) \\ \Psi^y(\alpha, 0) \\ \Psi^u(\alpha, 0) \end{pmatrix} + \epsilon^{11/2} \bar{R}_0(\alpha)$$

with $\|\bar{R}_0\|_{\mathfrak{H}^s} \leq C_I$ then the unique solution of (WW3) satisfies the estimate

$$\left\| \begin{pmatrix} z(\cdot, t) \\ y(\cdot, t) \\ u(\cdot, t) \end{pmatrix} - \begin{pmatrix} \Psi^z(\cdot, t) \\ \Psi^y(\cdot, t) \\ \Psi^u(\cdot, t) \end{pmatrix} \right\|_{\mathfrak{H}^s} \leq C_F \epsilon^{11/2}$$

for $t \in [0, T_0 \epsilon^{-3}]$. The constant C_F does not depend on ϵ.

3. THE OPERATOR $K(x, y)$ PART I: BASICS AND BASIC EXPANSIONS.

This is the first of two sections where we discuss the operator $K(x, y)$ which gives the relation between y_t and x_t in the water wave equation. Here we briefly discuss the origin of this operator and report expansions of K found in previous work. We also state some very basic facts about these expansions. In Section 5 we quote more complicated results and prove some new technical extensions needed for our purposes.

$K(x, y)$ (or, more precisely $-K(x, y)$) is sometimes called the Hilbert transform for the region $\Omega(t)$. Loosely, given a region Ω in the complex plane, and any function F which is analytic in Ω, the Hilbert transform for Ω, $H(\Omega)$, is a linear operator which relates the real and imaginary parts of F on the boundary of Ω. That is

$$\text{Im}(F)|_{\partial\Omega} = H(\Omega)\text{Re}(F)|_{\partial\Omega}.$$

For example if Ω were the lower half-plane, then the Hilbert transform would be the operator H, given by $\widehat{Hf} = i\,\text{sgn}(k)\widehat{f}$. (This particular operator H is also frequently called *the* Hilbert transform.) The nature of the operator depends greatly on the region begin studied. Unsurprisingly, the proof that such an operator exists is connected to the Riemann mapping theorem and to techniques for solving boundary value problems for Laplace's equation in the plane. In this problem, since the region $\Omega(t)$ is completely specified by the coordinate functions x and y, we denote the Hilbert transform by $H(\Omega(t)) = -K(x, y)$.

As we are considering a fluid which is incompressible and a flow which is irrotational, $x_t(\alpha, t) - iy_t(\alpha, t)$ is the value on the upper boundary $\Gamma(t)$, of an analytic function on $\Omega(t)$, $\omega(\alpha + i\beta) = v^x(\alpha, \beta) - iv^y(\alpha, \beta)$. Here (v^x, v^y) is the velocity field for the fluid in the whole region. Thus, given that $K(x, y)$ exists, we have

$$y_t = K(x, y)x_t.$$

Of course, the boundary of $\Omega(t)$ is not just $\Gamma(t)$, but also includes the bottom of the canal (*i.e.* where $\beta = -1$). As we do not have fluid flow through the bottom, we have $v^y(\alpha, -1) = 0$.

Under these conditions, $K(x,y)$ has been analyzed extensively by Craig in [9] and Schneider and Wayne in [31]. In particular Craig shows that $K(x,y)$ has the following expansion:

$$(8) \qquad K(x,y)u = K_0 u + K_1(x,y)u + S_2(x,y)u$$

where

$$\widehat{K_0 u}(k) = -i\tanh(k)\widehat{u}(k),$$

$$K_1(x,y)u = [x,K_0]\partial_\alpha u - (y + K_0(yK_0))\partial_\alpha u$$

and S_2 is quadratic in (x,y).

First of all we note that K_0 is a bounded operator from H^s to H^s since $\tanh(k)$ is a bounded function. That is

$$\|K_0 u\|_s \leq \|u\|_s.$$

The operator $L = K_0^{-1}\partial_\alpha$, which is well-defined as we discussed in the Section 2, will also be used frequently. L is not a bounded operator on H^s. It effectively takes one derivative. That is

$$(9) \qquad \|Lu\|_s \leq \|u\|_{s+1}.$$

On the other hand, L^{-1} replaces one derivative. That is, since

$$|\tanh(k)/k| \leq (1+k^2)^{-1/2}$$

we know

$$(10) \qquad \|L^{-1}u\|_s \leq \|u\|_{s-1}.$$

We will be considering functions which are of long wavelength. That is, functions of the form $f(\alpha) = F(\beta)$ where $\beta = \epsilon\alpha$. We define operators $K_{0,\epsilon}$ and L_ϵ via

$$K_{0,\epsilon}F(\beta) = K_0 f(\alpha)$$
$$L_\epsilon F(\beta) = Lf(\alpha).$$

Taking the Maclaurin series expansion for $\tanh(k)$ shows that formally

$$K_{0,\epsilon} = -\epsilon\partial_\beta - \frac{1}{3}\epsilon^3\partial_\beta^3 - \frac{2}{15}\epsilon^5\partial_\beta^5 + O(\epsilon^7).$$

Similarly L_ϵ and L_ϵ^{-1} have expansions in terms of derivatives

$$L_\epsilon = -1 + \frac{1}{3}\epsilon^2\partial_\beta^2 + \frac{1}{45}\epsilon^4\partial_\beta^4 + O(\epsilon^6)$$
$$L_\epsilon^{-1} = -1 - \frac{1}{3}\epsilon^2\partial_\beta^2 - \frac{2}{15}\epsilon^4\partial_\beta^4 + O(\epsilon^6).$$

We call such expansions of Fourier multiplier operators "long wave approximations". The rigorous connection between a long wave approximation and the original operator is given in the following Lemma, whose simple proof is contained in Section 7.

Lemma 1. *Suppose A and A_n are linear operators defined by $\widehat{Af}(k) = \widehat{A}(k)\widehat{f}(k)$, and $\widehat{A_nf}(k) = \widehat{A_n}(k)\widehat{f}(k)$ where $\widehat{A}(k)$ and $\widehat{A_n}(k)$ are complex valued functions. Also suppose that $|\widehat{A}(k) - \widehat{A_n}(k)| \leq C|k|^n$. (e.g. $\widehat{A_n}$ is a Taylor polynomial for $\widehat{A}$.) Then for $f \in H^{s+n}$ we have*

$$\|Af(\cdot) - A_nf(\cdot)\|_s \leq C\|\partial_\alpha^n f(\cdot)\|_s.$$

Moreover, if $f(\alpha)$ is of long wavelength form—that is if $f(\alpha) = F(\epsilon\alpha)$, with $F \in H^{s+n}$—then for $0 < \epsilon < 1$ there exists C independent of ϵ such that

$$\|Af(\cdot) - A_nf(\cdot)\|_s \leq C\epsilon^{n-1/2}\|F(\cdot)\|_{s+n}.$$

In [31], Schneider and Wayne show that the operator $K(x, y)$ does not depend on x *per se*, but rather on $z = K_0x$. We confuse the notation for the operators intentionally. They show the $K(z, y)$ has the following expansion

$$K(z, y)u = K_0u + K_1(z, y)u + S_2(z, y)u.$$

where

$$K_1(z, y)u = M_1(z)\partial_\alpha u - (y + K_0yK_0)\partial_\alpha u$$

with

$$\mathfrak{F}[M_1(z)v](k) = -\int \frac{\widehat{K_0}(k) - \widehat{K_0}(l)}{\widehat{K_0}(k - l)}\widehat{z}(k - l)\widehat{v}(l)dl.$$

S_2 is an operator which depends quadratically on z and y. Section 5 contains an analysis of these operators.

By using the hyperbolic trigonometric identity

(11)
$$\frac{\tanh(l) - \tanh(k)}{\tanh(l - k)} = 1 - \tanh(k)\tanh(l)$$

we can simplify the expression for K_1 to

(12)
$$\begin{aligned} K_1(z, y)u &= M_1(z + y)\partial_\alpha u \\ &= -(z + y + K_0(z + y)K_0)\partial_\alpha u \end{aligned}$$

Since we know K_0 is a bounded operator, it is clear that

(13)
$$\|M_1(z)v\|_s \leq \|zv\|_s.$$

4. THE DERIVATION

In this section we will derive the higher order correction to the KdV approximation. For technical reasons, it is most convenient to work with the water wave equation written in the form (WW3). Suppose that one is given the function $\Psi(\alpha, t) = (\Psi^z(\alpha, t), \Psi^y(\alpha, t), \Psi^u(\alpha, t))^{tr}$. The amount that this function fails to satisfy (WW3) is called the *residual* and is given by $\text{Res}[\Psi] = (\text{Res}_z, \text{Res}_y, \text{Res}_u)^{tr}$ with

$$\text{Res}_z = \partial_t\Psi^z - K_0\Psi^u$$

$$\text{Res}_y = \partial_t\Psi^y - K(\Psi^z, \Psi^y)\Psi^u$$

$$\text{Res}_u = \partial_t\Psi^u + \partial_\alpha\Psi^y\frac{1 + \partial_t^2\Psi^y}{1 + L\Psi^z}.$$

For a true solution, notice that $\text{Res}[\Psi]$ is identically zero.

Remark 3. *We will also consider the four dimensional system (WW4). If we let $\Psi^a = \partial_t \Psi^u$, then we have the additional Res function*

$$\begin{aligned}
Res_a =& \partial_t \Psi^a + \frac{\Psi^a \partial_\alpha \Psi^u + \partial_\alpha \partial_t (1 + \partial_t^2 \Psi^y) + \partial_t \Psi^y \partial_t^3 \Psi^y}{1 + L\Psi^z} \\
=& \partial_t Res_u + \frac{\partial_\alpha \Psi^u}{1 + L\Psi^z} Res_u.
\end{aligned}$$

The main goal when deriving modulation equations is to choose a system of equations such that solutions to this system yield a very small residual. This is different than (but connected to) showing that solutions to the modulation equations are close to true solutions for the original problem. This latter issue is precisely that answered by the main results, Theorem 2 and Proposition 3, and is discussed in Section 6. Here, we will perform a series of calculations on the residual and derive equations (KdV), (IW) and (LK). In this process we guarantee the smallness of the residual. While several of the steps will initially seem to have little mathematical justification (*i.e.* they are formal), once the calculation is completed it will be obvious that all steps are valid. For example, we will take

(14) $$\Psi^u = K_0^{-1} \partial_t \Psi^z.$$

With this choice

$$\text{Res}_z = 0,$$

which is small indeed! However, K_0^{-1} is not in general a well-defined operator. Nonetheless, when we eventually select Ψ^z, $K_0^{-1} \partial_t \Psi^z$ will make perfect sense.

We are looking for solutions which are small in amplitude and long in wavelength. So we let

$$\begin{aligned}
\Psi^z(\alpha, t) =& \epsilon^2 Z(\beta, \tau) \\
=& \epsilon^2 Z_1(\beta, \tau) + \epsilon^4 Z_2(\beta, \tau) + \epsilon^6 Z_3(\beta, \tau).
\end{aligned}$$

Recall $\beta = \epsilon\alpha$ and $\tau = \epsilon t$. We require $\text{Res}[\Psi]$ to be $O(\epsilon^{17/2})$. Loosely, we need three powers of ϵ more than the expected error of $O(\epsilon^{11/2})$ to account for the long times $(O(\epsilon^{-3}))$ over which our approximation will be a valid. See Schneider and Wayne [31] and Wayne and Wright [36].

Remark 4. *More specifically, if we wish to prove Theorem 2 in the space $\mathfrak{H}^s$ we will need*

$$\|Res_z\|_s \leq C\epsilon^{17/2}$$

$$\|Res_y\|_s \leq C\epsilon^{17/2}$$

$$\|Res_u\|_{s-1} \leq C\epsilon^{17/2}$$

$$\|Res_a\|_{s-1} \leq C\epsilon^{19/2}$$

for $0 \leq t \leq T_0\epsilon^{-3}$. Given the definition of Res_a, the final estimate will follow automatically from the estimate on Res_u.

We have already chosen Ψ^u in terms of Ψ^z. We will use first use the expression for Res_y to similarly determine Ψ^y in terms of Ψ^z. This is not as simple a matter because while K_0 commutes with ∂_t, the full operator $K(z, y)$ does not.

We have

$$\begin{aligned}
\text{Res}_y = & -\partial_t \Psi^y + K_0 \Psi^u + M_1(\Psi^z + \Psi^y)\partial_\alpha \Psi^u + S_2(\Psi^y, \Psi^z)\Psi^u \\
= & -\partial_t \Psi^y + \partial_t \Psi^z + M_1(\Psi^z + \Psi^y)K_0^{-1}\partial_\alpha \partial_t \Psi^z + S_2(\Psi^y, \Psi^z)\Psi^u.
\end{aligned}$$

Notice that in the above expression we can cancel the linear terms by taking $\Psi^y \sim \Psi^z$. More precisely, we set

$$\Psi^y(\alpha, t) = \epsilon^2 Z(\beta, \tau) + \epsilon^4 \Delta_1(\beta, \tau) + \epsilon^6 \Delta_2(\beta, \tau),$$

for as yet undetermined functions Δ_i. So

$$\begin{aligned}
\text{Res}_y = & -\epsilon^5 \partial_\tau \Delta_1 - \epsilon^7 \partial_\tau \Delta_2 + M_1(2\epsilon^2 Z)L_\epsilon \epsilon^3 \partial_\tau Z \\
& + M_1(\epsilon^4 \Delta_1 + \epsilon^6 \Delta_2)L_\epsilon \epsilon^3 \partial_\tau Z + S_2(\Psi^y, \Psi^z)\Psi^u \\
= & -\epsilon^5 \partial_\tau \Delta_1 - \epsilon^7 \partial_\tau \Delta_2 - 2\epsilon^5 Z L_\epsilon \partial_\tau Z - \epsilon^6 K_{0,\epsilon}\left(2Z\partial_\beta \partial_\tau Z\right) \\
& - \epsilon^7 \Delta_1 L_\epsilon \partial_\tau Z - \epsilon^8 K_{0,\epsilon}\left(\Delta_1 \partial_\beta \partial_\tau Z\right) + \epsilon^9 M_1(\Delta_2)L_\epsilon \partial_\tau Z \\
& + S_2(\Psi^y, \Psi^z)\Psi^u.
\end{aligned}$$

A number of the terms in Res_y are already $O(\epsilon^{17/2})$. By Lemma 1 we have the following estimate on $K_{0,\epsilon}$:

$$\begin{equation}
\|K_{0,\epsilon}F\|_s \leq \epsilon^{1/2}\|F\|_{s+1}. \tag{15}
\end{equation}$$

Thus terms containing $K_{0,\epsilon}$ can be considered to be a power of ϵ smaller than they appear (though this costs a derivative). On the other hand, K_0 is bounded so we have

$$\begin{equation}
\|K_{0,\epsilon}F\|_s \leq \epsilon^{-1/2}\|F\|_s. \tag{16}
\end{equation}$$

Thus we can use $K_{0,\epsilon}$ either as a bounded functional *or* to gain powers of ϵ, but not both. Notice that L_ϵ does not contribute any additional powers of ϵ in any case. We separate out all the terms that are already sufficiently small into error terms. That is

$$\begin{aligned}
\text{Res}_y = & -\epsilon^5 \partial_\tau \Delta_1 - \epsilon^7 \partial_\tau \Delta_2 - 2\epsilon^5 Z_1 L_\epsilon \partial_\tau Z_1 \\
& - 2\epsilon^7 Z_1 L_\epsilon \partial_\tau Z_2 - 2\epsilon^7 Z_2 L_\epsilon \partial_\tau Z_1 \\
& - \epsilon^6 K_{0,\epsilon}\left(2Z_1\partial_\beta \partial_\tau Z_1\right) - \epsilon^7 \Delta_1 L_\epsilon \partial_\tau Z_1 \\
& + E^y_{small} + E^y_{S_2}
\end{aligned}$$

with

$$E^y_{small} = 2\epsilon^9 Z_3 L_\epsilon \partial_\tau Z + 2\epsilon^9 (Z_1 + \epsilon^2 Z_2) L_\epsilon \partial_\tau Z_3 + 2\epsilon^9 Z_2 L_\epsilon \partial_\tau Z_2$$
$$- \epsilon^8 K_{0,\epsilon} \left((2Z_2 + 2\epsilon^2 Z_3) \partial_\beta \partial_\tau Z \right) - \epsilon^8 K_{0,\epsilon} \left(2Z_1 \partial_\beta \partial_\tau (Z_2 + \epsilon^2 Z_3) \right)$$
$$- \epsilon^9 \Delta_1 L_\epsilon \partial_\tau (Z_2 + \epsilon^2 Z_3) - \epsilon^8 K_{0,\epsilon} (\Delta_1 \partial_\beta \partial_\tau Z) + \epsilon^9 M_1(\Delta_2) L_\epsilon \partial_\tau Z$$
$$E^y_{S_2} = S_2(\Psi^y, \Psi^z) \Psi^u.$$

It is clear that E^y_{small} is $O(\epsilon^{17/2})$. That is

$$\|E^y_{small}\|_s \leq C \epsilon^{17/2}.$$

The constant C depends on various norms of the functions Z_i, $\partial_\tau Z_i$ and Δ_i. Specifically, chasing through the various terms in E^y_{small} and applying the estimates in (9), (13), (15) and (16), one can show that C depends on $\|Z_1\|_{s+1}$, $\|Z_2\|_{s+1}$, $\|Z_3\|_s$, $\|\partial_\tau Z_1\|_{s+2}$, $\|\partial_\tau Z_2\|_{s+2}$, $\|\partial_\tau Z_3\|_{s+1}$, $\|\Delta_1\|_{s+1}$ and $\|\Delta_2\|_s$.

The term $E^y_{S_2}$ is also $O(\epsilon^{17/2})$ though this is not as obvious. We prove this in Proposition 3 in Section 5. The proof of this relies strongly on the fact that we have taken Ψ^y and Ψ^z such that $\Psi^y - \Psi^z$ is $O(\epsilon^4)$. This causes a cancellation in S_2, which in turn makes this term small.

We now expand L_ϵ and $K_{0,\epsilon}$ in the remaining low order terms in Res_y to find

$$\text{Res}_y = -\epsilon^5 \partial_\tau \Delta_1 - \epsilon^7 \partial_\tau \Delta_2 + 2\epsilon^5 Z_1 \partial_\tau Z_1 + \frac{4}{3} \epsilon^7 Z_1 \partial_\tau \partial_\beta^2 Z_1$$
$$+ 2\epsilon^7 Z_1 \partial_\tau Z_2 + 2\epsilon^7 Z_2 \partial_\tau Z_1$$
$$+ 2\epsilon^7 \partial_\beta Z_1 \partial_\beta \partial_\tau Z_1 + \epsilon^7 \Delta_1 \partial_\tau Z_1$$
$$+ E^y_{small} + E^y_{S_2} + E^y_{lwa}$$

with

$$E^y_{lwa} = -2\epsilon^5 Z_1 (L_\epsilon + 1 - \frac{1}{3} \epsilon^2 \partial_\beta^2) \partial_\tau Z_1 - 2\epsilon^7 Z_1 (L_\epsilon + 1) \partial_\tau Z_2$$
$$- 2\epsilon^7 Z_2 (L_\epsilon + 1) \partial_\tau Z_1 - 2\epsilon^6 (K_{0,\epsilon} + \epsilon \partial_\beta)(Z_1 \partial_\beta \partial_\tau Z_1)$$
$$- \epsilon^7 \Delta_1 (L_\epsilon + 1) \partial_\tau Z_1.$$

Each term in E^y_{lwa} is $O(\epsilon^{17/2})$ by Lemma 1. That is

$$\|E^y_{lwa}\| \leq C \epsilon^{17/2}$$

where C depends on $\|Z_1\|_{s+3}, \|Z_2\|_s, \|\partial_\tau Z_1\|_{s+4}, \|\partial_\tau Z_2\|_{s+2}$ and $\|\Delta_1\|_s$. (The subscript "lwa" stands for "long wave approximation".)

The only $O(\epsilon^5)$ terms remaining in Res_y are

$$-\epsilon^5 \partial_\tau \Delta_1 + 2\epsilon^5 Z_1 \partial_\tau Z_1$$

which we remove by selecting

$$(17) \hspace{4cm} \Delta_1 = Z_1^2.$$

So

$$\mathrm{Res}_y = - \epsilon^7 \partial_\tau \Delta_2 + \frac{4}{3}\epsilon^7 Z_1 \partial_\tau \partial_\beta^2 Z_1$$
$$+ 2\epsilon^7 Z_1 \partial_\tau Z_2 + 2\epsilon^7 Z_2 \partial_\tau Z_1$$
$$+ 2\epsilon^7 \partial_\beta Z_1 \partial_\beta \partial_\tau Z_1 + \epsilon^7 Z_1^2 \partial_\tau Z_1$$
$$+ E_{small}^y + E_{S_2}^y + E_{lwa}^y.$$

The remaining $O(\epsilon^7)$ terms in Res_y are all perfect time derivatives with the exception of

$$\frac{4}{3}Z_1 \partial_\tau \partial_\beta^2 Z_1.$$

Notice, however, that

$$\partial_\tau \left(\frac{4}{3}Z_1 \partial_\beta^2 Z_1 - \frac{2}{3}(\partial_\tau Z_1)^2 \right)$$
$$= \frac{4}{3}Z_1 \partial_\tau \partial_\beta^2 Z_1 + \frac{4}{3}\partial_\tau Z_2 \left(\partial_\beta^2 Z_1 - \partial_\tau^2 Z_1 \right).$$

Given the form of the approximation in equation (1), it is not unreasonable to suspect that

(18) $$\partial_\beta^2 Z_1 - \partial_\tau^2 Z_1 \sim O(\epsilon^2).$$

We are now in a position to select Δ_2. Taking

(19)
$$\Delta_2 = (\partial_\beta Z_1)^2 + 2Z_1 Z_2 + \frac{1}{3}Z_1^3$$
$$+ \frac{4}{3}Z_1 \partial_\beta^2 Z_1 - \frac{2}{3}(\partial_\tau Z_1)^2$$

gives

$$\mathrm{Res}_y = E_{small}^y + E_{S_2}^y + E_{lwa}^y + E_{switch}^y$$

where

$$E_{switch}^y = \frac{4}{3}\epsilon^7 \partial_\tau Z_1 (\partial_\beta^2 Z_1 - \partial_\tau^2 Z_1).$$

Given that our assumption (18) is valid, we have shown that with our choices for Ψ^y and Ψ^u in terms of Ψ^z that $\mathrm{Res}_y = O(\epsilon^{17/2})$. More specifically, we have shown that if

$$\|\partial_\beta^2 Z_1 - \partial_\tau^2 Z_1\|_s \leq C\epsilon^{3/2},$$

then

$$\mathrm{Res}_y \leq C\epsilon^{17/2}$$

where the constant depends only on $\|Z_1\|_{s+3}$, $\|Z_2\|_{s+1}$, $\|Z_3\|_s$, $\|\partial_\tau Z_1\|_{s+4}$, $\|\partial_\tau Z_2\|_{s+2}$, $\|\partial_\tau Z_3\|_{s+1}$.

Now that we have computed Ψ^y and Ψ^u in terms of Ψ^z, we now turn our attention to determining Ψ^z by examining Res_u.

$$\mathrm{Res}_u = K_0^{-1}\partial_t^2 \Psi^z + \partial_\alpha \Psi^y \frac{1 + \partial_t^2 \Psi^y}{1 + L\Psi^z}.$$

We expand $(1 + L\Psi^z)^{-1}$ by the geometric series to find

$$\text{Res}_u = K_0^{-1}\partial_t^2\Psi^z + \partial_\alpha\Psi^y(1 + \partial_t^2\Psi^y)(1 - L\Psi^z + (L\Psi^z)^2) + E_1^u$$

where

$$E_{geo}^u = \partial_\alpha\Psi^y\frac{1 + \partial_t^2\Psi^y}{1 + L\Psi^z} - \partial_\alpha\Psi^y(1 + \partial_t^2\Psi^y)(1 - L\Psi^z + (L\Psi^z)^2).$$

Since Ψ^z is "small" this error term can be shown to be $O(\epsilon^{17/2})$. We have

Lemma 2. *Let $f \in H^{s+1}$, $s > 0$. Take ϵ_0 such that $\epsilon_0^2\|f\|_{L^\infty} \le 1/2$. Then for $0 < \epsilon < \epsilon_0$ we have that the function*

$$g(\epsilon x) = \frac{1}{1 + \epsilon^2 f(\epsilon x)} - 1 + \epsilon^2 f(\epsilon x)$$

satisfies $\|g(\epsilon\cdot)\|_s \le C\epsilon^{7/2}$ for C independent of ϵ.

Remark 5. *Under the same hypotheses as in Lemma 2, arguments similar to the proof of that Lemma show:*

- *$(1 + \epsilon^2 f(\epsilon x))^{-1} \in C^s$ and is bounded there independent of ϵ,*
- *$(1 + \epsilon^2 f(\epsilon x))^{-1} - 1 \in H^s$ and has norm there bounded by $C\epsilon^{3/2}$ for C independent of ϵ and*
- *$(1 + \epsilon^2 f(\epsilon x))^{-1} - 1 + \epsilon^2 f(\epsilon x) - \epsilon^4 f^2(\epsilon x) \in H^s$ and has norm there bounded by $C\epsilon^{11/2}$ for C independent of ϵ.*

Now, after substituting in from the definitions of Ψ^y and Ψ^z, we collect all the terms which are smaller than $O(\epsilon^{17/2})$ and find

$$\begin{aligned}
\text{Res}_u =\,&\epsilon^4 K_{0,\epsilon}^{-1}\partial_\tau^2(Z_1 + \epsilon^2 Z_2 + \epsilon^4 Z_3) + \epsilon^3\partial_\beta Z_1 - \epsilon^5\partial_\beta Z_1 L_\epsilon Z_1 \\
&+ \epsilon^5\partial_\beta\left(Z_2 + Z_1^2\right) + \epsilon^7\partial_\beta Z_1\left(\partial_\beta^2 Z_1 - L_\epsilon Z_2 + (L_\epsilon Z_1)^2\right) \\
&- \epsilon^7\partial_\beta(Z_2 + Z_1^2)L_\epsilon Z_1 + \epsilon^7\partial_\beta Z_3 \\
&+ \epsilon^7\partial_\beta\left((\partial_\beta Z_1)^2 + 2Z_1 Z_2 + \frac{1}{3}Z_1^3\right) \\
&+ \epsilon^7\partial_\beta\left(\frac{4}{3}Z_1\partial_\beta^2 Z_1 - \frac{2}{3}(\partial_\tau Z_1)^2\right) \\
&+ E_{geo}^u + E_{small}^u + E_{switch}^u.
\end{aligned}$$

We omit the exact expression for E_{small}^u because it is both lengthy and uninteresting. We have

(20) $$\|E_{small}^u\|_s \le C\epsilon^{17/2}$$

where the constant C depends on H^{s+1} norms of the functions Z_i and the H^s norms of $\partial_\tau^2 Z_i$. We have also replaced one instance of $\partial_\tau^2 Z_1$ with $\partial_\beta^2 Z_1$ (much as we did earlier), thus the term

$$E_{switch}^u = \epsilon^7\partial_\beta Z_1\left(\partial_\beta^2 Z_1 - \partial_\tau^2 Z_1\right).$$

We now define new functions W_i by $Z_i = L_\epsilon^{-1}W_i$. This seemingly mysterious (and sudden!) change of variables will seem less so if we remind

the reader that at the end of the day we wish model not z but rather the function x_α. Accordingly, if we approximate x_α by a function

$$\Psi^d = \epsilon^2 W_1 + \epsilon^4 W_2 + \epsilon^6 W_3,$$

then it is logical to take

$$\Psi^z = L^{-1}\Psi^d$$

and in the long wavelength limit we arrive at these functions W_i.

So we have

$$\begin{aligned}
\text{Res}_u =& \epsilon^3 \partial_\beta^{-1} \partial_\tau^2 (W_1 + \epsilon^2 W_2 + \epsilon^4 W_3) + \epsilon^2 K_{0,\epsilon} W_1 - \epsilon^4 W_1 K_{0,\epsilon} W_1 \\
& + \epsilon^4 K_{0,\epsilon} W_2 + \epsilon^5 \partial_\beta (L_\epsilon^{-1} W_1)^2 \\
& + \epsilon^2 K_{0,\epsilon} W_1 \left(\epsilon^3 K_{0,\epsilon} \partial_\beta W_1 - \epsilon^4 W_2 + \epsilon^4 W_1^2 \right) \\
& - (\epsilon^6 K_{0,\epsilon} W_2 + \epsilon^7 \partial_\beta (L_\epsilon^{-1} W_1)^2) W_1 + \epsilon^6 K_{0,\epsilon} W_3 \\
& + \epsilon^5 \partial_\beta \left((K_{0,\epsilon} W_1)^2 + 2\epsilon^2 L_\epsilon^{-1} W_1 L_\epsilon^{-1} W_2 + \frac{1}{3}\epsilon^2 (L_\epsilon^{-1} W_1)^3 \right) \\
& + \epsilon^7 \partial_\beta \left(\frac{4}{3} L_\epsilon^{-1} W_1 L_\epsilon^{-1} \partial_\beta^2 W_1 - \frac{2}{3}(L_\epsilon^{-1} \partial_\tau W_1)^2 \right) \\
& + E_{geo}^u + E_{small}^u + E_{switch}^u.
\end{aligned}$$

At this time, the presence of inverse β derivatives may seem problematic. Notice that each such precedes a time derivative. Once we select the functions W_i we will see that there can be an exchange between time and space derivatives, which will justify the instances of ∂_β^{-1}.

Now we replace $K_{0,\epsilon}$ and L_ϵ^{-1} by their long wave approximates and find

$$\begin{aligned}
\text{Res}_u =& \epsilon^3 \partial_\beta^{-1} \partial_\tau^2 (W_1 + \epsilon^2 W_2 + \epsilon^4 W_3) \\
& - \epsilon^2 (\epsilon \partial_\beta + \frac{1}{3}\epsilon^3 \partial_\beta^3 + \frac{2}{15}\epsilon^5 \partial_\beta^5) W_1 \\
& + \epsilon^4 W_1 (\epsilon \partial_\beta + \frac{1}{3}\epsilon^3 \partial_\beta^3) W_1 \\
& - \epsilon^4 (\epsilon \partial_\beta + \frac{1}{3}\epsilon^3 \partial_\beta^3) W_2 + \epsilon^5 \partial_\beta (W_1)^2 + \frac{2}{3}\epsilon^7 \partial_\beta (W_1 \partial_\beta^2 W_1) \\
& - \epsilon^7 \partial_\beta W_1 \left(-\partial_\beta^2 W_1 - W_2 + W_1^2 \right) \\
& - \epsilon^7 W_1 (-\partial_\beta W_2 + \partial_\beta (W_1)^2) - \epsilon^7 \partial_\beta W_3 \\
& + \epsilon^7 \partial_\beta \left((\partial_\beta W_1)^2 + 2W_1 W_2 - \frac{1}{3} W_1^3 \right) \\
& + \epsilon^7 \partial_\beta \left(\frac{4}{3} W_1 \partial_\beta^2 W_1 - \frac{2}{3}(\partial_\tau W_1)^2 \right) \\
& + E_{geo}^u + E_{small}^u + E_{switch}^u + E_{lwa}^u.
\end{aligned}$$

The error made by the long wave approximations is denoted E_{lwa}^u. By Lemma 1 we have

$$\|E_{lwa}^u\|_s \le C\epsilon^{17/2}$$

where C depends on $\|W_1\|_{s+7}$, $\|W_2\|_{s+5}$, $\|W_3\|_{s+3}$, $\|\partial_\tau W_1\|_{s+2}$.

Now we organize the above as

$$
\begin{aligned}
\operatorname{Res}_y =\ & \epsilon^3 \partial_\beta^{-1} \partial_\tau^2 W_1 - \epsilon^3 \partial_\beta W_1 \\
& - \epsilon^5 \frac{1}{3} \partial_\beta^3 W_1 + \epsilon^5 \frac{3}{2} \partial_\beta (W_1)^2 \\
& + \epsilon^5 \partial_\beta^{-1} \partial_\tau^2 W_2 - \epsilon^5 \partial_\beta W_2 \\
& - \epsilon^7 \frac{1}{3} \partial_\beta^3 W_2 + \epsilon^7 3 \partial_\beta (W_1 W_2) \\
& - \frac{2}{15} \epsilon^7 \partial_\beta^5 W_1 + \epsilon^7 \frac{1}{3} W_1 \partial_\beta^3 W_1 \\
& + \epsilon^7 2 \partial_\beta (W_1 \partial_\beta^2 W_1) + \epsilon^7 \frac{3}{2} \partial_\beta (\partial_\beta W_1)^2 \\
& - \epsilon^7 \frac{4}{3} \partial_\beta (W_1)^3 - \epsilon^7 \frac{2}{3} \partial_\beta (\partial_\tau W_1)^2 \\
& + \epsilon^7 \partial_\beta^{-1} \partial_\tau^2 W_3 - \epsilon^7 \partial_\beta W_3 \\
& + E_{geo}^u + E_{small}^u + E_{switch}^u + E_{lwa}^u.
\end{aligned}
$$

(21)

The term on the first line of right hand side looks formally like an inverse derivative of a wave equation:

$$
\partial_\beta^{-1} \partial_\tau^2 W_1 - \partial_\beta W_1 = \partial_\beta^{-1} \left(\partial_\tau^2 W_1 - \partial_\beta^2 W_1 \right).
$$

We cancel this term (to lowest order) by taking W_1 of the form

$$
W_1(\beta, \tau) = -U(\beta - \tau, \epsilon^2 \tau) - V(\beta + \tau, \epsilon^2 \tau).
$$

Recall $\beta_\pm = \beta \pm \tau$ and $T = \epsilon^2 \tau$. The "minus" signs may seem arbitrary, but are included at this stage so that they agree with previous work in the area. Noting that the third line looks very much like the first, we also set

$$
W_2(\beta, \tau) = -F(\beta - \tau, \epsilon^2 \tau) - G(\beta + \tau, \epsilon^2 \tau) - P(\beta, \tau).
$$

These choices for W_1 and W_2 are precisely those described heuristically in the Introduction.

The first three lines in (21) become

$$
\begin{aligned}
& \epsilon^5 \left(2 \partial_T U + \frac{1}{3} \partial_{\beta_-}^3 U + \frac{3}{2} \partial_{\beta_-} U^2 \right) \\
+ & \epsilon^5 \left(-2 \partial_T V + \frac{1}{3} \partial_{\beta_+}^3 V + \frac{3}{2} \partial_{\beta_+} V^2 \right) \\
+ & \epsilon^5 \left(\partial_\beta^{-1} \left(\partial_\beta^2 P - \partial_\tau^2 P \right) + 3 \partial_\beta (UV) \right) \\
+ & \epsilon^7 \left(2 \partial_T F - 2 \partial_T G - \partial_\beta^{-1} \left(\partial_T^2 U + \partial_T^2 V \right) \right) \\
- & \epsilon^9 \partial_\beta^{-1} \left(\partial_T^2 F + \partial_T^2 G \right).
\end{aligned}
$$

We cancel everything multiplied by ϵ^5 by taking

$$-2\partial_T U = \frac{1}{3}\partial_{\beta_-}^3 U + \frac{3}{2}\partial_{\beta_-} U^2$$

$$2\partial_T V = \frac{1}{3}\partial_{\beta_+}^3 V + \frac{3}{2}\partial_{\beta_+} V^2$$

$$\partial_\tau^2 P - \partial_\beta^2 P = 3\partial_\beta^2(UV)$$

which are precisely equations (KdV) and (IW). By Proposition 1 we know the solutions to these equations are well-behaved over the long time scales.

Given that the functions U and V have been chosen to solve (KdV), one computes that

$$\partial_T^2 U = \partial_{\beta_-}\left(\frac{1}{36}\partial_{\beta_-}^5 U + \frac{9}{4}U^2\partial_{\beta_-} U + \frac{1}{2}U\partial_{\beta_-}^3 U + \frac{3}{4}\partial_{\beta_-} U\partial_{\beta_-}^2 U\right)$$

$$\partial_T^2 V = \partial_{\beta_+}\left(\frac{1}{36}\partial_{\beta_+}^5 V + \frac{9}{4}V^2\partial_{\beta_+} V + \frac{1}{2}V\partial_{\beta_+}^3 V + \frac{3}{4}\partial_{\beta_+} V\partial_{\beta_+}^2 V\right).$$

Thus the term $\partial_\beta^{-1}\left(\partial_T^2 U + \partial_T^2 V\right)$ is perfectly well-defined. For brevity, we will continue to write these terms with the inverse derivatives instead of in the longer form above.

Moreover, now we can put more precise estimates on E^y_{switch} and E^u_{switch}. In particular, since each time derivative for solutions to KdV equations count for three space derivatives, we have

$$\|\partial_\beta^2 W_1 - \partial_\tau^2 W_1\|_s \leq C\epsilon^{3/2}$$

where C depends on $\|W_1\|_{s+6}$.

Recall from Fact 1 and Proposition 1 that solutions to (IW) can be rewritten as

$$P(\beta,\tau) = P^+(\beta,\tau) + P^-(\beta,\tau)$$
$$= \varphi^+(\beta_+,T) + \varphi^-(\beta_-,T).$$

The functions $\varphi^\pm$ are rapidly decaying. We make this decomposition so that every remaining term in (21):

- will be a unidirectional term which is rapidly decaying;
- will be a product of two such terms which are moving in opposite directions;
- or will include a derivative of W_3.

That is

$$\text{Res}_y = \epsilon^7 \left(2\partial_T F + \frac{1}{3}\partial_{\beta_-}^3 F + 3\partial_{\beta_-}(UF) + J^- \right)$$
$$+ \epsilon^7 \left(-2\partial_T G + \frac{1}{3}\partial_{\beta_+}^3 G + 3\partial_{\beta_+}(VG) + J^+ \right)$$
$$+ \epsilon^7 \left(\partial_\beta^{-1}\left(\partial_\tau^2 W_3 - \partial_\beta^2 W_3\right) + J^s \right)$$
$$- \epsilon^9 \partial_\beta^{-1}\left(\partial_T^2 F + \partial_T^2 G\right)$$
$$+ E_{geo}^u + E_{small}^u + E_{switch}^u + E_{lwa}^u + E_{time}^u.$$

where

$$
\begin{aligned}
J^- = \;& 3\partial_{\beta_-}(U\varphi^-) + 4U^2\partial_{\beta_-}U + \frac{7}{3}U\partial_{\beta_-}^3 U \\
& + \frac{11}{3}\partial_{\beta_-}U\partial_{\beta_-}^2 U + \frac{2}{15}\partial_{\beta_-}^5 U + \frac{1}{3}\partial_{\beta_-}^3\varphi^- \\
& - \partial_{\beta_-}^{-1}\partial_T^2 U \\
J^+ = \;& 3\partial_{\beta_+}(V\varphi^+) + 4V^2\partial_{\beta_+}V + 7/3 V\partial_{\beta_+}^3 V \\
& + \frac{11}{3}\partial_{\beta_+}V\partial_{\beta_+}^2 V + \frac{2}{15}\partial_{\beta_+}^5 V + \frac{1}{3}\partial_{\beta_+}^3\varphi^+ \\
& - \partial_{\beta_+}^{-1}\partial_T^2 V \\
J^s = \;& \partial_\beta\left(U(3G + 3\varphi^+ + 4V^2 + \frac{7}{3}\partial_{\beta_+}^2 V) \right) \\
& + \partial_\beta\left(V(3F + 3\varphi^- + 4U^2 + \frac{7}{3}\partial_{\beta_-}^2 U) \right) \\
& + 4\partial_\beta\left(\partial_{\beta_-}U\partial_{\beta_+}V \right).
\end{aligned}
$$

(22)

and

$$E_{time}^u = + \epsilon^9\frac{4}{3}\partial_\beta\left(\left(\partial_{\beta_-}U - \partial_{\beta_+}V\right)\left(\partial_T U + \partial_T V\right) \right)$$
$$- \epsilon^{11}\frac{2}{3}\partial_\beta\left((\partial_T U + \partial_T V)^2 \right).$$

Notice that $J^\pm = J^\pm(\beta_\pm, T)$. E_{time}^u (so called because each term in it contains some sort of time derivative) is clearly $O(\epsilon^{17/2})$. That is

$$\|E_{time}^u\|_s \leq C\epsilon^{17/2}.$$

The constant above depends on $\|U\|_{s+4}$ and $\|V\|_{s+4}$.

The term $\epsilon^9\partial_\beta^{-1}\left(\partial_T^2 F + \partial_T^2 G\right)$ is not included in E_{time}^u for the following reason. In a moment, when we select the equations F and G solve, a consequence will be that there will be terms in $\partial_T^2 F$ and $\partial_T^2 G$ which are $O(\epsilon^{-2})$.

By taking

$$-2\partial_T F = \frac{1}{3}\partial_{\beta_-}(UF) + \frac{3}{2}\partial_{\beta_-}^3 F + J^-$$

$$2\partial_T G = \frac{1}{3}\partial_{\beta_+}(VG) + \frac{3}{2}\partial_{\beta_+}^3 G + J^+$$

we cancel nearly all the terms which are not in the various E^u terms. These are the linearized KdV equations (LK) discussed in the Introduction. Proposition 1 guarantees that the solutions are well-behaved. We are left with

$$\mathrm{Res}_y = + \epsilon^7 \left(\partial_\beta^{-1} \left(\partial_\tau^2 W_3 - \partial_\beta^2 W_3 \right) + J^s \right)$$
$$- \epsilon^9 \partial_\beta^{-1} \left(\partial_T^2 F + \partial_T^2 G \right)$$
$$+ E^u_{geo} + E^u_{small} + E^u_{switch} + E^u_{lwa} + E^u_{time}.$$

Now we consider the terms in $\partial_\beta^{-1} \left(\partial_T^2 F + \partial_T^2 G \right)$. Notice that

$$-\partial_{\beta_-}^{-1}\partial_T^2 F = \partial_{\beta_-}^{-1}\partial_T(\frac{1}{6}\partial_\beta^3 F + \frac{3}{2}\partial_\beta(UF) + \frac{1}{2}J^-)$$
$$= \frac{1}{6}\partial_{\beta_-}^2 \partial_T F + \frac{3}{2}\partial_T(UF) + \frac{1}{2}\partial_{\beta_-}^{-1}\partial_T J^-.$$

J^- contains the term $3\partial_{\beta_-}(U\varphi^-) + \frac{1}{3}\partial_{\beta_-}^3 \varphi^-$. From the definition of φ^- we know

$$\partial_T \varphi^- = -\epsilon^{-2}\frac{3}{2}\partial_\beta(UV).$$

So we have

$$\frac{1}{2}\partial_{\beta_-}^{-1}\partial_T(3\partial_{\beta_-}(U\varphi^-) + \frac{1}{3}\partial_{\beta_-}^3 \varphi^-)$$
$$= \frac{3}{2}\varphi^- \partial_T U + \frac{3}{2}U\partial_T \varphi^- + \frac{1}{6}\partial_{\beta_-}^2 \partial_T \varphi^-$$
$$= \frac{3}{2}\varphi^- \partial_T U - \frac{9}{4}\epsilon^{-2}U\partial_\beta(UV) + \frac{1}{4}\epsilon^{-2}\partial_\beta^3(UV).$$

We treat $\partial_T^2 G$ in the same fashion. So we can write

$$-\epsilon^9 \partial_\beta^{-1} \left(\partial_T^2 F + \partial_T^2 G \right) = E^u_{F,G} - \epsilon^7 \left(\frac{9}{4}(U+V)\partial_\beta(UV) + \frac{1}{2}\partial_\beta^3(UV) \right).$$

By construction $E^u_{F,G}$ satisfies the estimate

$$\|E^u_{F,G}\|_s \le C\epsilon^{17/2}$$

where C depends on $\|U\|_{s+7}$, $\|V\|_{s+7}$, $\|F\|_{s+5}$, and $\|G\|_{s+5}$.

We have

$$\mathrm{Res}_y = + \epsilon^7 \left(\partial_\beta^{-1} \left(\partial_\tau^2 W_3 - \partial_\beta^2 W_3 \right) + J^s \right)$$
$$- \epsilon^7 \left(\frac{9}{4}(U+V)\partial_\beta(UV) + \frac{1}{2}\partial_\beta^3(UV) \right)$$
$$+ E^u_{geo} + E^u_{small} + E^u_{switch} + E^u_{lwa} + E^u_{time} + E^u_{F,G}.$$

By selecting

$$(23) \qquad \partial_\tau^2 W_3 - \partial_\beta^2 W_3 = \partial_\beta \left(\frac{9}{4}(U+V)\partial_\beta(UV) + \frac{1}{2}\partial_\beta^3(UV) - J^s \right)$$

the gun goes off and we cancel all remaining $O(\epsilon^7)$ terms. Thus

$$\text{Res}_y = E_{geo}^u + E_{small}^u + E_{switch}^u + E_{lwa}^u + E_{time}^u + E_{F,G}^u.$$

Each of the E^u is $O(\epsilon^{17/2})$.

Unlike the previous equations (KdV), (IW) and (LK), Proposition 1 does not tell us that the solutions to (23) are controllable. Nonetheless, equation (23) is an inhomogeneous wave equation where the inhomogeneity consists entirely of terms which are products of left and right moving rapidly decaying functions. From Wayne and Wright [36] we have the following Lemma:

Lemma 3. *Suppose*

$$\partial_\tau u \pm \partial_\beta u = l(\beta + \tau, \epsilon^2\tau)r(\beta - \tau, \epsilon^2\tau), \quad u(X,0) = 0.$$

with $\|l(\cdot, T)\|_{H^s(2)} \leq C$ *and* $\|r(\cdot, T)\|_{H^s(2)} \leq C$ *for* $T \in [0, T_0]$, *then*

$$\|u(\beta, \tau)\|_s \leq C$$

for all $\tau \in [0, T_0\epsilon^{-2}]$. *The constant* C *is uniform in* ϵ.

Thus W_3 will remain $O(1)$.

Remark 6. *If we are in the situation in which Proposition 1 applies, we see that least regular part in the driving term are* $\partial_\beta^2(UG)$ *and* $\partial_\beta^2(VF)$, *which are in* $H^{\sigma-6}(2)$. *Thus, by this Lemma we have that* $W_3 \in H^{\sigma-5}$ *for all times of interest.*

At this time we have derived the modulation equations and shown the residual is small. The only remaining order of business in this section is to determine how smooth the solutions to our modulation equations need to be in order for $\text{Res}[\Psi]$ to be appropriately regular. This may seem to be a fairly tiresome task, but fortunately the least regular terms in all of the sundry E functions come from only one term—E_{small}^u! This is because E_{small}^u contains many time derivatives.

We need to control Res_a in H^{s-1}. For this we need $\partial_t\text{Res}_u \in H^{s-1}$, which in turn implies that we must have $\partial_t E_{small}^u \in H^{s-1}$. Recalling equation (20), we see that this will require $\partial_\tau^3 Z_2 \in H^{s-1}$, or rather (since L_ϵ^{-1} saves a derivative) $\partial_\tau^3 W_2 \in H^{s-2}$. For this, we need $\partial_T^3 F$ and $\partial_T^3 G$ in H^{s-2}. Given that F solves (LK) where J^+ contains the terms $\partial_{\beta_-}^5 U$, $\partial_{\beta_-}^3 \varphi^-$, one sees that $\partial_T^3 F$ will include the terms $\partial_{\beta_-}^9 F$, $\partial_{\beta_-}^{11} U$ and $\partial_{\beta_-}^9 \varphi^-$. So $\|\partial_T^3 F\|_{s-2}$ is controlled by the H^{s+9} norms of U and V, and the H^{s+7} norms of φ^-, F and G. The analogous result is true for $\partial_T^3 G$. We also need $\partial_\tau^3 W_3 \in H^{s-2}$. Since W_3 solves (23), we require $W_3 \in H^{s+1}$.

In summary we have the following Proposition.

Proposition 2. *Take Ψ^d as in (7), with U, V, F, G, P and W_3 solving their respective equations. Let*

$$\Psi^z = L^{-1}\Psi^d$$
$$\Psi^y = \Psi^z + \epsilon^4\Delta_1 + \epsilon^6\Delta_2$$
$$\Psi^u = \partial_\alpha^{-1}\partial_t\Psi^d$$
$$\Psi^a = \partial_\alpha^{-1}\partial_t^2\Psi^d$$

with Δ_1 and Δ_2 as in (17) and (19), and form $Res[\Psi]$ as in (4). Then

$$\|Res_z\|_s \leq C\epsilon^{17/2}$$
$$\|Res_y\|_s \leq C\epsilon^{17/2}$$
$$\|Res_u\|_{s-1} \leq C\epsilon^{17/2}$$
$$\|Res_a\|_{s-1} \leq C\epsilon^{19/2}.$$

where C is a constant which depends on $\|U\|_{s+9}$, $\|V\|_{s+9}$, $\|P\|_{s+7}$, $\|F\|_{s+7}$, $\|G\|_{s+7}$ and $\|W_3\|_{s+1}$. The estimate (4) holds as long as these quantities remain bounded. The constant C does not depend on ϵ.

In light of Proposition 1 and Remark 6, we see that if in we take the initial conditions for U and V to satisfy (2) with $\sigma \geq s + 7$, that $\|U\|_{s+9}$, $\|V\|_{s+9}$, $\|P\|_{s+7}$, $\|F\|_{s+7}$, $\|G\|_{s+7}$ and $\|W_3\|_{s+1}$ are all $O(1)$ for $t \in [0, T_0\epsilon^{-3}]$. And so we move on.

5. THE OPERATOR $K(x,y)$ PART II: ESTIMATES AND EXTENSIONS

In this section we will describe a few more estimates related to $K(x,y)$. All such estimates are either smoothing estimates or ones which show that certain terms are small in the long wavelength setting.

First, since $1 + \widehat{K_0}^2(k)$ goes to zero exponentially fast as $|k| \to \infty$, the operator $1 + K_0^2$ is smoothing. That is, for all $s \geq 0$

$$\|(1 + K_0^2)u\|_s \leq C\|u\|_{L^2}.$$

Also, commutators involving K_0 are smoothing. We quote the following Lemma from [31].

Lemma 4. *Let $r \geq 0$, $q > 1/2$, and $0 \leq p \leq q$. Then there exists a $C > 0$ such that*

$$\|[f, K_0]g\|_r \leq C\|f\|_{r+p}\|g\|_{q-p}.$$

Proof. See Lemma 3.12 on p 1498 of [31]. $\qquad\square$

Schneider and Wayne show that $K_1(z,y)$ is a smoothing operator.

Lemma 5. *For $r \geq 0$, $q \geq 1/2$ and $0 \leq p \leq q$, there is C such that*

(24) $$\|K_1(z,y)u\|_r \leq C\left(\|z\|_{r+p} + \|y\|_{r+p}\right)\|u\|_{q-p}.$$

Proof. See Corollary 3.13 on p 1499 of [31]. $\qquad\square$

If we let $S_1(z, y) = K(z, y) - K_0$, we also have the following estimates from [31]:

Lemma 6. *Fix $s \geq 4$. If the free surface is sufficiently smooth, then for $j = 1, 2$ we have:*

* $$\|S_j(z, y)u\|_s \leq C \left(\|z\|_s^j + \|y\|_s^j \right) \|u\|_3,$$

    that is, S_j is a smoothing operator,

* $$\|\partial_\alpha(S_j(z, y)u)\|_s \leq C \left(\|z\|_s^j + \|y\|_s^j \right) \|u\|_3,$$

    that is, $\partial_\alpha S_j$ is a smoothing operator,

* $$\|[\partial_t, S_j]u\| \leq C \left(\|z\|_s^j + \|y\|_s^j \right) \|u\|_3,$$

    that is, $[\partial_t, S_j]$ is a smoothing operator and this operator can be bounded independently of $\partial_t u$ and

* $$\|[\partial_t^2, S_j]u\| \leq C \left(\|z\|_s^j + \|y\|_s^j \right) (\|u\|_4 + \|\partial_t u\|_4),$$

    that is, $[\partial_t^2, S_j]$ is a smoothing operator and this operator can be bounded independently of $\partial_t^2 u$.

Proof. In [31], see Lemmas 3.14, 3.15 and Corollary 3.16 on *pp* 1500, 1506 and 1507 respectively. $\qquad\square$

We will also need the following propositions concerning the behavior of the remainder terms S_1 and S_2. The first of these says that more or less the remainder S_2 is negligible for the sort of scalings we are considering. That is to say, the term $E_{S_2}^y$ in the Section 4 is very small.

Proposition 3. *Fix $s > 5/2$. Suppose $z = \epsilon^2 Z(\epsilon\alpha)$, $y = \epsilon^2 Y(\epsilon\alpha)$ and $f = \epsilon^2 F(\epsilon\alpha)$, with $Z, Y, F \in H^{s+1}(2)$. Moreover, assume $z - y = \epsilon^4 \Delta(\epsilon\alpha)$ with $\Delta \in H^{s+1}(2)$. Then there exist ϵ_0 such that for $\epsilon \in [0, \epsilon_0]$ there is a constant C independent of ϵ such that:*

$$\|S_2(z, y)f\|_s \leq C\epsilon^{17/2}.$$

The second is a technical version of the mean value theorem as applied to the operator S_1.

Proposition 4. *Suppose $z(\alpha, t) = \epsilon^2 Z(\epsilon(\alpha \pm t), \epsilon^3 t)$, $y(\alpha, t) = \epsilon^2 Y(\epsilon(\alpha \pm t), \epsilon^3 t)$, $u(\alpha, t) = \epsilon^2 U(\epsilon(\alpha \pm t), \epsilon^3 t)$ and $f(\alpha, t) = \epsilon^2 F(\epsilon(\alpha \pm t), \epsilon^3 t)$ with $Z, Y, U, F \in H^s(2)$ for $t \in [0, T_0\epsilon^{-3}]$. Also suppose $R^z(\alpha, t)$, $R^y(\alpha, t)$ and $R^u(\alpha, t) \in H^s$ for the same time interval. Then*

$$\|S_1(z(\cdot) + \epsilon^{11/2} R^z(\cdot), y(\cdot) + \epsilon^{11/2} R^y(\cdot))f(\cdot) - S_1(z(\cdot), y(\cdot))f(\cdot)\|_s \leq C\epsilon^{17/2}$$

for $t \in [0, T_0\epsilon^{-3}]$.

Proof. **for Proposition 3**:

First, notice $x(\alpha) = \int_0^\alpha Lz(a)da = \epsilon X(\epsilon\alpha)$. We know that X is in L^∞ by the following Lemma.

Lemma 7. *Suppose $f(\alpha) = \epsilon^2 F(\epsilon\alpha)$ with $F \in H^s(2)$. Then for all α*

$$\left| \int_0^\alpha f(a)da \right| \leq C\epsilon\|F\|_{H^s(2)}$$

Proof. See Section 7. $\qquad\qquad\square$

Let $\Phi(\tilde{x}, y) = (\Phi_1, \Phi_2)$ be the analytic map which takes $\Omega(t)$ to

$$P^- = \{(\xi, \gamma)|\gamma \in [-1, 0]\} \,.$$

That such a map exists and is analytic is guaranteed by the Riemann mapping theorem. Let

$$h(\alpha) = \Phi_1(\tilde{x}(\alpha), y(\alpha))$$

and $Qf = f \circ h$. From [31], we know that

(25) $$K(x, y)f(\alpha) = Q \circ K_0 \circ Q^{-1}f(\alpha).$$

We can derive a very useful implicit formula for h^{-1} as follows. The function $\Phi^{-1}(\xi, \gamma)$ is analytic on P^-, thus it satisfies the Cauchy-Riemann equations. If we set

$$\Phi^{-1}(\xi, \gamma) = (\xi + u_1(\xi, \gamma)) + i\,(\gamma - v_1(\xi, \gamma))$$

and notice that Φ^{-1} sends the bottom and top of P^- to the bottom and top of $\Omega(t)$ respectively, we see that we have the following system

$$\partial_\xi u_1 + \partial_\gamma v_1 = 0,$$
$$\partial_\gamma u_1 - \partial_\xi v_1 = 0,$$
$$v_1(\xi, -1) = 0, \quad v_1(\xi, 0) = \eta(\xi)$$

where $\eta(\xi) = y(h^{-1}(\xi))$. One can solve this system with the use of Fourier transforms relatively simply. If one does so, one finds that

$$u_1(\xi, 0) = -\int_0^\xi \eta(\xi_1)d\xi_1 - M\eta(\xi)$$

where M is the pseudo-differential operator given by

$$\widehat{M\eta}(k) = \frac{k\cosh(k) - \sinh(k)}{ik\sinh(k)}\widehat{\eta}(k).$$

Notice that to lowest order, M is $C\partial_\alpha$.

Now, notice that $h^{-1}(\xi) = \tilde{x}^{-1}(u(\xi, 0))$ and so we have an implicit equation for h^{-1}.

(26)
$$h^{-1}(\xi) = \tilde{x}^{-1}\left(\xi - \int_0^\xi \eta(\xi_1)d\xi_1 - M\eta(\xi)\right)$$
$$= \tilde{x}^{-1}\left(\xi - \int_0^\xi y(h^{-1}(\xi_1))d\xi_1 - M(y \circ h^{-1})(\xi)\right).$$

Remark 7. *In Schneider and Wayne [31], on p 1494, they make a minor error in calculating this same function. As a result, they claim the above representation gives an explicit formula for h^{-1}. Our correction here changes nothing about subsequent steps in their proofs.*

Since, $\tilde{x} = \alpha + \epsilon X(\epsilon\alpha)$ where X is well-behaved, we can expect a similar form for $\tilde{x}^{-1}$.

Lemma 8. *Suppose $f(\alpha) = \alpha + g(\alpha)$ with $\|g\|_{C^2} \leq 1/2$. Then*

$$f^{-1}(\xi) = \xi - g(\xi) + g(\xi)g'(\xi) + E$$

where $E = O(\|g\|_{W^{2,\infty}}^3)$. More specifically

$$E \leq C\left(\|g'\|_{L^\infty}^2\|g\|_{L^\infty} + \|g\|_{L^\infty}^2\|g''\|_{L^\infty}\right).$$

In particular, notice that if $g(\alpha) = \epsilon G(\epsilon\alpha)$ this means $E = O(\epsilon^5)$.

Proof. See Section 7. □

We apply this Lemma to $\tilde{x}$, and find that

$$\tilde{x}^{-1}(\xi) = \xi - \epsilon X(\epsilon\xi) + \epsilon^3 X(\epsilon\xi)\partial_\beta X(\epsilon\xi) + O(\epsilon^5).$$

Combining this with (26) we can determine $h^{-1}(\xi)$ (and therefore h) in terms of x and y to any order we wish. To lowest order we see that

(27) $$h^{-1}(\xi) = \xi + O(\epsilon)$$

So now we have

$$
\begin{aligned}
h^{-1}(\xi) =& \xi - \int_0^\xi y(h^{-1}(\xi_1))d\xi_1 - M(y \circ h^{-1})(\xi) \\
& - \epsilon X\left(\epsilon\left(\xi - \int_0^\xi y(h^{-1}(\xi_1))d\xi_1 - M(y \circ h^{-1})(\xi)\right)\right) \\
& + O(\epsilon^3) \\
=& \xi - \int_0^\xi \epsilon^2 Y(\epsilon h^{-1}(\xi_1))d\xi_1 - M(\epsilon^2 Y \circ \epsilon h^{-1})(\xi) \\
& - \epsilon X\left(\epsilon\left(\xi - \int_0^\xi \epsilon^2 Y(\epsilon h^{-1}(\xi_1))d\xi_1 - M(\epsilon^2 Y \circ \epsilon h^{-1})(\xi)\right)\right) \\
& + O(\epsilon^3).
\end{aligned}
$$

If we insert (27) into the above and expand we have

$$h^{-1}(\xi) = \xi - \epsilon X(\epsilon\xi) - \int_0^\xi \epsilon^2 Y(\epsilon\xi_1))d\xi_1 - \epsilon^2 M(Y(\epsilon\cdot))(\xi) + O(\epsilon^3).$$

One can continue in this manner and determine the next order terms in the expansion of h^{-1}. If we let

$$\epsilon G_1(\epsilon\alpha) = -\epsilon X(\epsilon\alpha) - \int_0^\alpha \epsilon^2 Y(\epsilon\alpha_1))d\alpha_1 - \epsilon^2 M(Y(\epsilon\cdot))(\alpha).$$

the expansion is

$$h^{-1}(\xi) = \xi + \epsilon G_1(\epsilon\xi) + \epsilon^3 B_1(\epsilon\xi) + O(\epsilon^5)$$

where

$$B_1(\xi) = \int_0^\xi \epsilon G_1(\epsilon a)\partial_\beta \epsilon^3 Y(\epsilon a)da$$
$$+ M(\epsilon^3 G_1(\epsilon\cdot)\partial_\beta Y(\epsilon\cdot))(\xi) + \epsilon^3 G_1(\epsilon\xi)\partial_\beta X(\epsilon\xi).$$

Notice that since M is $C\partial_\alpha$ to lowest order, $-\epsilon^2 M(Y(\epsilon\cdot))(\alpha)$ is $O(\epsilon^3)$. Moreover, by hypothesis, we have $\epsilon^2 Z(\epsilon\alpha) - \epsilon^2 Y(\epsilon\alpha) = \epsilon^4\Delta(\epsilon\alpha)$. Thus

$$-\epsilon X(\epsilon\alpha) - \int_0^\alpha \epsilon^2 Y(\epsilon a)da = -\int_0^\alpha \left(\epsilon^2\partial_\beta X(\epsilon a) + \epsilon^2 Y(\epsilon a)\right)da$$
$$= -\int_0^\alpha \left(\epsilon^2 L(Z(\epsilon\cdot))(a) + \epsilon^2 Y(\epsilon a)\right)da$$
$$= \int_0^\alpha \left(\epsilon^2 Z(\epsilon a) - \epsilon^2 Y(\epsilon a)\right)da + O(\epsilon^3)$$
$$= \int_0^\alpha \left(\epsilon^4\Delta(\epsilon a)\right)da + O(\epsilon^3)$$
$$= O(\epsilon^3).$$

That is, ϵG_1 is really $O(\epsilon^3)$! This cancelation is the crucial step in this proof. Since ϵG_1 appears in each term in B_1, we have shown

$$h^{-1}(\xi) = \xi + \epsilon^3 G(\epsilon\xi) + O(\epsilon^5)$$

with $\epsilon^3 G = \epsilon G_1$. We appeal to Lemma 8 again, and we have

$$h(\alpha) = \alpha - \epsilon^3 G(\epsilon\alpha) + O(\epsilon^5).$$

Now that we have particularly good estimates on h and h^{-1}, we can begin our discussion of K in earnest. For notational simplicity, we will let

$$h(\alpha) = \alpha + g_1(\alpha)$$
$$h^{-1}(\xi) = \xi + g_2(\xi)$$

If we let

$$\tilde{f} = Q^{-1}f$$

we can make the following formal approximation using Taylor's theorem,

$$K(x,y)f(\alpha) = Q \circ K_0\tilde{f}(\alpha)$$
$$= K_0\tilde{f}(h(\alpha))$$
$$= K_0\tilde{f}(\alpha + g_1(\alpha))$$
$$= K_0\tilde{f}(\alpha) + g_1(\alpha)K_0\partial_\alpha\tilde{f}(\alpha) + h.o.t..$$

Also by Taylor's theorem,

$$\tilde{f}(\alpha) = f(\alpha) + g_2(\alpha)\partial_\alpha f(\alpha) + h.o.t..$$

Putting these together we have

$$K(x,y)f(\alpha) = K_0 f(\alpha) + g_1(\alpha)K_0 f'(\alpha) + K_0(g_2 f')(\alpha) + h.o.t..$$

So let

$$E_1 f = K_0 f(\alpha) + g_1(\alpha)K_0 f'(\alpha) + K_0(g_2 f')(\alpha)$$
$$E_2 f = K(x,y)f - E_1 f.$$

We prove Proposition 3 if we can prove

- $\|E_1 f - K_0 f - K_1(x,y)f\|_s \le C\epsilon^{17/2}$, and
- $\|E_2 f\|_s \le C\epsilon^{17/2}$.

Let us deal with $E_2 f$ first. We can rewrite $E_2 f$ as:

$$E_2 f = E_2^1 f + E_2^2 f + E_2^3 f$$

with

$$E_2^1 f = K(x,y)f - K_0\tilde{f} - g_1 K_0 \partial_\alpha \tilde{f}$$
$$E_2^2 f = K_0\tilde{f} - K_0 f - K_0(g_2 \partial_\alpha f)$$
$$E_2^3 f = g_1 K_0 \partial_\alpha \tilde{f} - g_1 K_0 \partial_\alpha f$$

As our approximation for K was determined by an application of Taylor's theorem, we need to prove a Lemma which shows that this formal step can be made rigorous, at least for functions in the weighted Sobolev spaces.

Lemma 9. *Suppose $F \in H^s(n)$, $s > 1/2$, $n > 1/2$. Then for all $C_0 > 0$ there exists ϵ_0 such that for $\epsilon \in [0, \epsilon_0]$ there is a constant C independent of ϵ such that:*

$$\left(\int_{|\alpha| > C_0 \epsilon^{-3}} |F(\epsilon\alpha)|^2 d\alpha \right)^{1/2} \le C\epsilon^{2n-3/2}.$$

Moreover, for $1 \le j \le s$ we have:

$$\left(\int_{|\alpha| > C_0 \epsilon^{-3}} |\partial_\alpha^j F(\epsilon\alpha)|^2 d\alpha \right)^{1/2} \le C\epsilon^{2n-3/2+j}.$$

Proof. See Section 7. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 8. *If instead we are considering*

$$\left(\int_{|\alpha| > C_0 \epsilon^{-3}} |F(\epsilon(\alpha \pm t), \epsilon^3 t)|^2 d\alpha \right)^{1/2}$$

with $F(\cdot, T) \in H^s(n)$ for $T \in [0, T_0]$, we can maintain the same bound as above by taking $C_0 \ge 2T_0$.

We can use the above Lemma to prove a version of Taylor's theorem.

Lemma 10. *Suppose $F \in H^s(2)$, $s > 5/2$ and $g \in L^\infty$. Then*

$$\|F(\epsilon \cdot + \epsilon^2 g(\epsilon \cdot)) - F(\epsilon \cdot)\|_{L^2} \le C\epsilon^{3/2}.$$

Proof. By Lemma 9 we have

$$\|F(\epsilon \cdot + \epsilon^2 g(\epsilon \cdot)) - F(\epsilon \cdot)\|^2$$

$$= \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon\alpha + \epsilon^2 g(\epsilon\alpha)) - F(\epsilon\alpha)|^2 d\alpha$$

$$+ \int_{|\alpha| \geq \epsilon^{-3}} |F(\epsilon\alpha + \epsilon^2 g(\epsilon\alpha)) - F(\epsilon\alpha)|^2 d\alpha$$

$$\leq \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon\alpha + \epsilon^2 g(\epsilon\alpha)) - F(\epsilon\alpha)|^2 d\alpha + C\epsilon^5.$$

Now, we add and subtract $\epsilon^2 g(\epsilon\alpha)F'(\epsilon\alpha)$ in the remaining integral,

$$\int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon\alpha + \epsilon^2 g(\epsilon\alpha)) - F(\epsilon\alpha)|^2 d\alpha$$

$$\leq \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon\alpha + \epsilon^2 g(\epsilon\alpha)) - F(\epsilon\alpha) - \epsilon^2 g(\epsilon\alpha)F'(\epsilon\alpha)|^2 d\alpha$$

$$+ \int_{|\alpha| \leq \epsilon^{-3}} |\epsilon^2 g(\epsilon\alpha)F'(\epsilon\alpha)|^2 d\alpha$$

$$\leq \int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon\alpha + \epsilon^2 g(\epsilon\alpha)) - F(\epsilon\alpha) - \epsilon^2 g(\epsilon\alpha)F'(\epsilon\alpha)|^2 d\alpha$$

$$+ \epsilon^3 \|g(\cdot)\|_{L^\infty}^2 \|F'(\cdot)\|_{L^2}^2.$$

We naively bound the above integral and apply the mean value theorem. That is,

$$\int_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon\alpha + \epsilon^2 g(\epsilon\alpha)) - F(\epsilon\alpha) - \epsilon^2 g(\epsilon\alpha)F'(\epsilon\alpha)|^2 d\alpha$$

$$\leq C\epsilon^{-3} \sup_{|\alpha| \leq \epsilon^{-3}} |F(\epsilon\alpha + \epsilon^2 g(\epsilon\alpha)) - F(\epsilon\alpha) - \epsilon^2 g(\epsilon\alpha)F'(\epsilon\alpha)|^2$$

$$\leq C\epsilon^{-3} \sup_{|\alpha| \leq \epsilon^{-3}} |\epsilon^4 g^2(\epsilon\alpha)F''(\epsilon\alpha^*)|^2$$

$$\leq C\epsilon^5 \|g\|_{L^\infty}^4 \|F''\|_{L^\infty}^2.$$

With this, we have proven the Lemma. $\qquad\square$

Remark 9. *With this general technique we are also able to show that*

$$\|\partial_\alpha^j (F(\epsilon \cdot + \epsilon^2 g(\epsilon \cdot)) - F(\epsilon \cdot))\|_{L^2} \leq C\epsilon^{3/2+j},$$

$$\|\partial_\alpha^j \left(F(\epsilon \cdot + \epsilon^2 g(\epsilon \cdot)) - F(\epsilon \cdot) - \epsilon^2 g(\epsilon \cdot)F'(\cdot) \right)\|_{L^2} \leq C\epsilon^{7/2+j},$$

$$\|\partial_\alpha^j \left(F(\epsilon \cdot + \epsilon^2 g(\epsilon \cdot)) \right)$$
$$- \partial_\alpha^j \left(F(\epsilon \cdot) + \epsilon^2 g(\epsilon \cdot)F'(\cdot) + 1/2\epsilon^4 g^2(\epsilon \cdot)F''(\epsilon \cdot) \right)\|_{L^2} \leq C\epsilon^{11/2+j}$$

and so on.

Now we will be able to control E_2^1. We can control the other functions in precisely the same fashion. Since f is long wavelength and of rapid decay, so is $\tilde{f}$. And thus we can use Lemma 10. In what follows, $\epsilon^3 \tilde{F}(\epsilon\alpha) = K_0 \tilde{f}(\alpha)$. (The extra ϵ comes from the long wave approximation of K_0.)

$$
\begin{aligned}
&\|E_2^1 f(\cdot)\|_s \\
=& \|K_0 \tilde{f}(\cdot + g_1(\cdot)) - K_0 \tilde{f}(\cdot) - g_1(\cdot) K_0 \partial_\alpha \tilde{f}(\cdot)\|_s \\
\leq & C\epsilon^3 \|\tilde{F}(\epsilon \cdot + \epsilon^4 G(\epsilon\cdot)) - \tilde{F}(\epsilon\cdot) - \epsilon^4 G(\epsilon\cdot)\partial_\beta \tilde{F}(\epsilon\cdot)\|_s \\
\leq & C\epsilon^{17/2}
\end{aligned}
$$

Now we turn our attention to $E_1 f - K_0 f - K_1(x, y)f$. A routine calculation shows that this is equal to

$$
\begin{aligned}
&- \epsilon^3 [G(\epsilon\alpha), K_0]\epsilon^3 \partial_\beta F(\epsilon\alpha) - K_1(\epsilon X, \epsilon^2 Y)\epsilon^2 F(\epsilon\alpha) \\
&+ (g_1 + \epsilon^3 G)K_0\epsilon^3 \partial_\beta F(\epsilon\alpha) + K_0 \left((g_2 - \epsilon^3 G)\epsilon^3 \partial_\beta F(\epsilon\cdot)\right)(\alpha)
\end{aligned}
$$

Now, $g_1 + \epsilon^3 G$ and $g_2 - \epsilon^3 G$ are $O(\epsilon^5)$, so the second line above can be bounded by $C\epsilon^{17/2}$, if we make use of K_0's long wavelength approximation. Moreover, we claim that the first line is identically zero. Let

$$
\begin{aligned}
b(\alpha) &= - \int_0^\alpha \epsilon^2 Y(\epsilon\alpha_1)d\alpha_1 - \epsilon^2 M(Y(\epsilon\cdot))(\alpha) \\
&= - \int_0^\alpha y(\alpha_1)d\alpha_1 - My(\alpha).
\end{aligned}
$$

Thus $\epsilon^3 G(\epsilon\alpha) = b(\alpha) - x(\alpha)$. So

$$
\epsilon^3 [G(\epsilon\alpha), K_0]\epsilon^3 \partial_\beta F(\epsilon\alpha) = -[x(\alpha), K_0]\partial_\alpha f(\alpha) + [b(\alpha), K_0]\partial_\alpha f(\alpha).
$$

Taking the Fourier transform of the second term we have

$$
\begin{aligned}
=& \mathfrak{F}\left([b(\alpha), K_0]\partial_\alpha f(\alpha)\right)(k) \\
=& \int \left(\widehat{K_0}(l) - \widehat{K_0}(k)\right)\widehat{b}(l - k)il\widehat{f}(l)dl \\
=& \int \frac{\widehat{K_0}(l) - \widehat{K_0}(k)}{i(l - k)}i(l - k)\widehat{b}(l - k)il\widehat{f}(l)dl \\
=& \int \frac{\widehat{K_0}(l) - \widehat{K_0}(k)}{i(l - k)}\widehat{\partial_\alpha b}(l - k)il\widehat{f}(l)dl
\end{aligned}
$$

Now notice that $\partial_\alpha b = Ly$, so the above becomes

$$\int \frac{\widehat{K_0}(l) - \widehat{K_0}(k)}{i(l-k)} \widehat{Ly}(l-k)il\widehat{f}(l)dl$$

$$= \int \frac{\widehat{K_0}(l) - \widehat{K_0}(k)}{\widehat{K_0}(l-k)} \widehat{y}(l-k)il\widehat{f}(l)dl$$

$$= \int (1 + \widehat{K_0}(k)\widehat{K_0}(l))\widehat{y}(l-k)il\widehat{f}(l)dl$$

$$= \mathfrak{F}\left((y + K_0yK_0)\partial_\alpha f\right)(k).$$

where we have used the trigonometric identity (11) from Section 4.

That is

$$\epsilon^3[G(\epsilon\alpha), K_0]\epsilon^3\partial_\beta F(\epsilon\alpha)$$

$$= -[x(\alpha), K_0]\partial_\alpha f(\alpha) + ((y + K_0yK_0)\partial_\alpha f)(k)$$

$$= -K_1(x,y)f$$

and so we are done with the proof of Proposition 3.

$$\square$$

Proof. **for Proposition 4**:

Let h be as in the above subsection and h_2 be the analogous function for the configuration $(z_2, y_2) = (z + \epsilon^{11/2}R^z, y + \epsilon^{11/2}R^y)$. We also define the function $\tilde{x}_2 = \alpha + x_2$ by $\partial_\alpha x_2 = Lz_2$. Unlike in the previous Lemma, here the time dependence of the functions is important. And so we determine x and x_2 by integrating in both space and time. That is,

$$\tilde{x}(\alpha, t) = \alpha + \epsilon\chi(\epsilon\alpha, \epsilon t)$$

$$\tilde{x}_2(\alpha, t) = \alpha + \epsilon\chi(\epsilon\alpha, \epsilon t) + \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho(\alpha, t).$$

with

$$\epsilon\chi(\epsilon\alpha, \epsilon t) = \left(\int_0^t u(0,s)\,ds + \int_0^\alpha Lz(w,t)\,dw\right)$$

$$\epsilon^{5/2}E(t) = \epsilon^{11/2}\int_0^t R^u(0,s)\,ds$$

$$\epsilon^{11/2}\rho = \int_0^\alpha LR^z(w,t)\,dw$$

The functions satisfy the following estimates for all $t \in [0, T_0\epsilon^{-3}]$.

$$|\epsilon\chi(\epsilon\alpha, \epsilon t)| \le C\epsilon$$

$$|\epsilon^{5/2}E(t)| \le C\epsilon^{5/2}$$

$$|\epsilon^{11/2}\rho(\alpha, t)| \le C\sqrt{|\alpha|}\|R^z\|_{H^s}$$

The first estimate follows from similar estimates in the previous Lemma, the second is the naive bound and the final follows from following simple fact:

Fact 2. *If $p(\alpha) = \int_0^\alpha r(a)da$, where $r \in L^2$, then*

$$|p(\alpha)| \leq \sqrt{|\alpha|}\|r\|_{L^2}.$$

In what follows we will make strong use of the fact that $E(t)$ does not depend on α.

Using the same techniques as were used in proving Lemma 8, one can show that:

$$\tilde{x}_2^{-1}(\xi, t) = \tilde{x}^{-1}(\xi, t) - \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho_2(\xi)$$

where $|\rho_2(\alpha)| \leq \sqrt{|\alpha|}\|R^z\|_{L^2}$. This sort of estimate carries over to the functions h. That is

$$h_2^{-1}(\xi, t) = h^{-1}(\xi, t) - \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho_3(\xi, t)$$

$$h_2(\alpha, t) = h(\alpha, t) + \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho_4(\alpha, t)$$

where $|\rho_3(\alpha, t)|, |\rho_4(\alpha, t)| \leq \sqrt{|\alpha|}(\|R^z\|_{L^2} + \|R^y\|_{L^2})$ over the long time scale.

Now define $Q_2 f = f \circ h_2$. So

$$S_1(z + \epsilon^{11/2}R^z, y + \epsilon^{11/2})f - S_1(z, y)f$$

$$= Q_2 \circ K_0 \circ Q_2^{-1}f - Q \circ K_0 \circ Q^{-1}f$$

$$= Q \circ (Q^{-1} \circ Q_2 \circ K_0 \circ Q_2^{-1} \circ Q - K_0) \circ Q^{-1}f$$

Since Q and its inverse are bounded operators from H^s to H^s, we need only prove the estimate for the operator

$$\tilde{Q} \circ K_0 \circ \tilde{Q}^{-1} - K_0$$

where $\tilde{Q} = Q^{-1} \circ Q_2$. Notice that $\tilde{Q} \circ K_0 \circ \tilde{Q}^{-1}$ is the Hilbert operator K for a domain with the "h" function given by $\check{h}(\alpha) = h_2(h^{-1}(\alpha))$. Moreover, from the above calculations for h and h_2 we have,

$$h_2(h^{-1}(\alpha)) = \alpha + \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho_5(\alpha, t)$$

with ρ_5 satisfying the same type of estimates as ρ_4.

At this point we can make an appeal to Lemma 3.14 on p 1500 of [31]. In this lemma they prove that $\|S_1(z, y)f\|_s \leq C(\|z\|_s + \|y\|_s)\|f\|_3$. In the course of their proof, they show that if $h(\alpha) = \alpha + g(\alpha)$ then

$$\|Q \circ K_0 \circ Q^{-1}f(\cdot) - K_0 f(\cdot)\|_s \leq C\|\partial_\alpha g\|_{s-1}\|\partial_\alpha f\|_2.$$

(See the inequalities in Cases I-IV on pp 1501-1506.) So if we set $\tilde{g} = \epsilon^{5/2}E(t) + \epsilon^{11/2}\rho_5(\alpha, t)$, we see that taking a spatial derivative leaves us with

$$\partial_\alpha \tilde{g} = O(\epsilon^{11/2}).$$

And so, if we keep in mind that $f(\alpha, t) = \epsilon^2 F(\epsilon(\alpha \pm t, \epsilon^3 t)$,

$$\|\tilde{Q} \circ K_0 \circ \tilde{Q}^{-1}f(\cdot) - K_0 f(\cdot)\|_s \leq C\|\partial_\alpha \tilde{g}\|_{s-1}\|\partial_\alpha f\|_2$$

$$\leq C\epsilon^{17/2}$$

This completes the proof of Proposition 4.   □

6. The error estimates

In this section we prove that the approximation is rigorous. That is we will prove Theorem 2. We will be working with the three and four dimensional formulations of the water wave problem (equations (WW3) and (WW4)). From [31], we know that for initial data of the type we are considering, solutions to these equations exist over the long times we are considering. If (z, y, u) is a solution to (WW3), let:

$$
\begin{aligned}
z(\alpha, t) &= \Psi^z(\alpha, t) + \epsilon^{11/2} R^z(\alpha, t) \\
(28) \qquad y(\alpha, t) &= \Psi^y(\alpha, t) + \epsilon^{11/2} R^y(\alpha, t) \\
u(\alpha, t) &= \Psi^u(\alpha, t) + \epsilon^{11/2} R^u(\alpha, t)
\end{aligned}
$$

with the functions Ψ defined as above. We call R^z, R^y and R^u "error" functions and we denote $\bar{R} = (R^z, R^y, R^u)$. Our goal will be to show that $\bar{R}$ remain $O(1)$ in $\mathfrak{H}^s = H^s \times H^s \times H^{s-1/2}$ over the long time scale. If we can do this, then we will have proven the main theorem. The first step will be to determine the equations which these functions satisfy. Loosely, we want to be able to write for each of the error functions an evolution equation of the form

$$
\partial_t R = \text{quasilinear} + \text{small and smooth}.
$$

We will at times go to great lengths to achieve this!

Clearly,

$$
\partial_t R^z = K_0 R^u.
$$

Finding the equations for R^y and R^u is a bit more complex. First we focus on R^y. Substituting from (28) into $\partial_t y = K(z, y)u$, we have

$$
\begin{aligned}
&\partial_t R^y \\
&= \epsilon^{-11/2} \left(K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) \left(\Psi^u + \epsilon^{11/2} R^u \right) - \partial_t \Psi^y \right) \\
&= \ K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) R^u \\
&\quad + \epsilon^{-11/2} \left(K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y)\Psi^u - \partial_t \Psi^y \right) \\
&= \ K_0 R^u + M_1(\Psi^z)\partial_\alpha R^u - (\Psi^y + K_0(\Psi^y K_0))\partial_\alpha R^u + N^y
\end{aligned}
$$

where

$$
\begin{aligned}
&N^y \\
&= \epsilon^{-11/2}\mathrm{Res}_y \\
&\quad + \epsilon^{-11/2} \left(\left(S_1(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) - S_1(\Psi^z, \Psi^y) \right) \Psi^u \right) \\
&\quad + \left(K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) - K_0 - K_1(\Psi^z, \Psi^y) \right) R^u.
\end{aligned}
$$

We claim that N^y is "small". That is, we have:

Lemma 11. *For all $C_R > 0$, there exists ϵ_0 such that for all $\epsilon \in (0, \epsilon_0)$ and t such that $\sup_{0 \le t' \le t} \|\bar{R}(\cdot, t')\|_{\mathfrak{H}^s} \le C_R$ we have:*

$$\|N^y\|_s \le C \left(\epsilon^3 + \epsilon^3 \|\bar{R}\|_{\mathfrak{H}^s} + \epsilon^{11/2} \|\bar{R}\|_{\mathfrak{H}^s}^2 \right).$$

Proof. First we remark that the approximating functions Ψ, and their derivatives are all bounded over the long time scales. Thus, we will not be keeping track of the dependence of the norm of N^y on the norms of these functions. By Proposition 2, we know that $\|\epsilon^{-11/2} \mathrm{Res}_y\|_s \le C\epsilon^3$.

We can bound

$$\epsilon^{-11/2} \left(S_1(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) - S_1(\Psi^z, \Psi^y) \right) \Psi^u$$

by Lemma 4.

Finally,

$$\left\| \left(K(\Psi^z + \epsilon^{11/2} R^z, \Psi^y + \epsilon^{11/2} R^y) - K_0 - K_1(\Psi^z, \Psi^y) \right) R^u \right\|_s$$

$$\le C \left(\epsilon^3 \|\bar{R}\|_{\mathfrak{H}^s} + \epsilon^{11/2} \|\bar{R}\|_{\mathfrak{H}^s}^2 \right)$$

by the estimates on K and its expansions which we saw in Sections 3 and 5 (in particular Lemma 5 and Lemma 6). $\qquad \square$

Now we discuss R^u. We know that

(29) $$\partial_t u (1 + Lz) + \partial_\alpha y (1 + \partial_t^2 y) = 0.$$

We would like an evolution type equation for R^u. Notice that since $\partial_t y = K(z, y)u$, there is a "hidden" $\partial_t u$ in the term $\partial_t^2 y$. Recall that the commutator $[\partial_t, S_1(z, y)]u$ can be bounded independently of $\partial_t u$ (see Lemma 6 in Section 5). Therefore, we can rewrite the above as

$$(1 + Lz + \partial_\alpha y K(z, y)) \partial_t u + \partial_\alpha y (1 + [\partial_t, S_1(z, y)]u) = 0.$$

Substituting in for u from (28) the above becomes

$$\begin{aligned}
0 = {} & (1 + Lz + \partial_\alpha y K(z, y)) \partial_t \epsilon^{11/2} R^u \\
& + (1 + Lz + \partial_\alpha y K(z, y)) \partial_t \epsilon^2 \Psi^u \\
& + \partial_\alpha y (1 + [\partial_t, S_1(z, y)] \Psi^u) \\
& + \partial_\alpha y \left([\partial_t, S_1(z, y)] \epsilon^{11/2} R^u \right)
\end{aligned}$$

or rather

$$\begin{aligned}
0 = {} & (1 + Lz + \partial_\alpha y K(z, y)) \partial_t \epsilon^{11/2} R^u \\
& + (1 + Lz) \partial_t \epsilon^2 \Psi^u \\
& + \partial_\alpha y (1 + \partial_t (K(z, y) \Psi^u)) \\
& + \partial_\alpha y \left([\partial_t, S_1(z, y)] \epsilon^{11/2} R^u \right).
\end{aligned}$$

We rearrange this a bit, and break up y and z.

$$\begin{aligned}
0 = &(1 + Lz + \partial_\alpha y K(z,y)) \, \partial_t R^u + \partial_\alpha R^y \\
&+ LR^z \partial_t \Psi^u + \partial_\alpha R^y \partial_t (K(z,y)\Psi^u) + \partial_\alpha y [\partial_t, S_1(z,y)] R^u \\
&+ \epsilon^{-11/2} \left((1 + L\Psi^z) \, \partial_t \epsilon^2 \Psi^u + \partial_\alpha \Psi^y (1 + \partial_t (K(z,y)\Psi^u)) \right).
\end{aligned}$$

The operator

$$A(z,y) = (1 + Lz + \partial_\alpha y K(z,y))$$

is invertible since $K(z,y)$ is a bounded operator on H^s, provided z and y are small (which they are). Moreover, we can approximate A^{-1} *via* the Neumann series. Thus the above equation can be rewritten as

$$\partial_t R^u = -(1 - \epsilon^2 W_1)\partial_\alpha R^y + N^u$$

where $N^u = N_1^u + N_2^u$ and

$$\begin{aligned}
N_1^u = &- \epsilon^{-11/2} A^{-1} \left((1 + L\Psi^z) \, \partial_t \epsilon^2 \Psi^u + \partial_\alpha \Psi^y (1 + \partial_t (K(z,y)\Psi^u)) \right) \\
N_2^u = &- A^{-1} \left(LR^z \partial_t \Psi^u + \partial_\alpha R^y \partial_t (K(z,y)\Psi^u) + \partial_\alpha y [\partial_t, S_1(z,y)] R^u \right) \\
&+ \left(-A^{-1} + (1 - \epsilon^2 W_1) \right) \partial_\alpha R^y.
\end{aligned}$$

Lemma 12. *For all $C_R > 0$, there exists ϵ_0 such that for all $\epsilon \in (0, \epsilon_0)$ and t such that $\sup_{0 \le t' \le t} \|\bar{R}(\cdot, t')\|_{\mathfrak{H}^s} \le C_R$ we have*

$$\|N^u\|_{s-1} \le C \left(\epsilon^3 + \epsilon^3 \|\bar{R}\|_{\mathfrak{H}^s} + \epsilon^{11/2} \|\bar{R}\|_{\mathfrak{H}^s}^2 \right).$$

Proof. First we point out this estimate is in H^{s-1}. The loss of regularity here is easily seen. Both LR^z and $\partial_\alpha R^y$ explicitly appear in N_2^u, and are not smoothed by any operators. Thus, losing this derivative is unavoidable. In fact, it is easy to see that the above estimates holds for N_2^u by noting that A^{-1}, K and $[\partial_t, S_1]$ are bounded operators.

Bounding N_1^u is also easily done once we recognize that this term is almost exactly $\epsilon^{-11/2}\mathrm{Res}_u$. We have

$$\begin{aligned}
&\epsilon^{-11/2} A^{-1} \left((1 + L\Psi^z) \, \partial_t \epsilon^2 \Psi^u + \partial_\alpha \Psi^y (1 + \partial_t (K(z,y)\Psi^u)) \right) \\
=&\epsilon^{-11/2} A^{-1} \left((1 + L\Psi^z) \, \partial_t \epsilon^2 \Psi^u + \partial_\alpha \Psi^y (1 + \partial_t (K(\Psi^z, \Psi^y)\Psi^u)) \right) \\
&+ N_3^u \\
=&\epsilon^{-11/2} \left(\partial_t \epsilon^2 \Psi^u + \partial_\alpha \Psi^y \frac{1 + \partial_t (K(\Psi^z, \Psi^y)\Psi^u)}{1 + L\Psi^z} \right) + N_3^u + N_4^u \\
=&\epsilon^{-11/2} \left(\partial_t \epsilon^2 \Psi^u + \partial_\alpha \Psi^y \frac{1 + \partial_t^2 (\Psi^y)}{1 + L\Psi^z} \right) + N_3^u + N_4^u + N_5^u \\
=&\epsilon^{-11/2}\mathrm{Res}_u + N_3^u + N_4^u + N_5^u
\end{aligned}$$

where

$$N_3^u = \epsilon^{-11/2} A^{-1} \left(\partial_\alpha \Psi^y \partial_t \left(K(z,y) \Psi^u - K(\Psi^z, \Psi^y) \Psi^u \right) \right)$$

$$N_4^u = \epsilon^{-11/2} (A^{-1} - (1 + L\Psi^z)^{-1})$$
$$\times \left((1 + L\Psi^z) \partial_t \epsilon^2 \Psi^u + \partial_\alpha \Psi^y \left(1 + \partial_t (K(\Psi^z, \Psi^y) \Psi^u) \right) \right)$$

$$N_5^u = \epsilon^{-11/2} \partial_\alpha \Psi^y \frac{\partial_t (\mathrm{Res}_y)}{1 + L\Psi^z}.$$

We bound N_3^u using mean value theorem arguments entirely analogous to those used when bounding N^y. To bound N_4^y, one observes that

$$(1 + L\Psi^z) \partial_t \epsilon^2 \Psi^u + \partial_\alpha \Psi^y \left(1 + \partial_t (K(\Psi^z, \Psi^y) \Psi^u) \right)$$

is very nearly Res_u and is thus $O(\epsilon^{17/2})$. N_5^u is clearly small, as it contains $\partial_t \mathrm{Res}_y$. This completes the proof. $\qquad\square$

We need to make analogous calculations for the four dimensional system. Let

$$a(\alpha, t) = \epsilon^3 \Psi^a(\epsilon\alpha, \epsilon t) + \epsilon^{11/2} R^a(\alpha, t)$$

and $\bar{R}_e = (R^z, R^y, R^u, R^a)$. This extended set of error functions lives in $\mathfrak{H}_e^s = H^s \times H^s \times H^{s-1/2} \times H^{s-1}$.

It is easy to see that

$$\partial_t R^u = R^a$$

but more difficult to determine the evolution of R^a. We begin by taking a time derivative of (29).

$$(30) \qquad \partial_t^2 u(1 + Lz) + \partial_t u \partial_\alpha u + \partial_\alpha \partial_t y(1 + \partial_t^2 y) + \partial_\alpha y \partial_t^3 y = 0.$$

Letting,

$$I = \partial_t^2 u(1 + Lz) + \partial_\alpha y \partial_t^3 y$$
$$II = \partial_\alpha \partial_t y(1 + \partial_t^2 y) + \partial_t u \partial_\alpha u$$

(30) is $I + II = 0$.

Manipulations very similar to those carried out in determing $\partial_t R^u$, show that

$$I = A(z,y) \partial_t \epsilon^{11/2} R^a + \partial_\alpha y [\partial_t^2, S_1(z,y)] \epsilon^{11/2} R^u$$
$$+ (1 + Lz) \partial_t \epsilon^3 \Psi^a + \partial_\alpha y \partial_t^2 (K(z,y) \Psi^u).$$

For II, we have

$$II = (1 + \partial_t^2 y) \partial_\alpha \partial_t \epsilon^{11/2} R^y + \partial_t u \partial_\alpha \epsilon^{11/2} R^u$$
$$+ (1 + \partial_t^2 y) \partial_\alpha \partial_t \Psi^y + \partial_t u \partial_\alpha \Psi^u$$
$$= \epsilon^{11/2} \left((1 + \partial_t^2 y) \partial_\alpha (K_0 R^u + K_1(\Psi^z, \Psi^y) R^u + N^y) + \partial_t u \partial_\alpha R^u \right)$$
$$+ (1 + \partial_t^2 y) \partial_\alpha \partial_t \Psi^y + \partial_t u \partial_\alpha \Psi^u$$
$$= \epsilon^{11/2} \left((1 + \partial_t^2 y - \partial_t u K_0 \cdot) K_0 \partial_\alpha R^u + \partial_\alpha K_1(\Psi^z, \Psi^y) R^u \right)$$
$$+ (1 + \partial_t^2 y) \partial_\alpha \partial_t \Psi^y + \partial_t u \partial_\alpha \Psi^u + \epsilon^{11/2} B_{II}$$

where

$$B_{II} = \partial_t u(1 + K_0^2)\partial_\alpha R^u + \partial_t^2 y \partial_\alpha(K_1(\Psi^z, \Psi^y)R^u + N^y) + \partial_\alpha N^y.$$

Noting that $\partial_t^2 y = K_0 a + [\partial_t, S_1]u + S_1 a$, we see that that B_{II} is smooth in the error functions, and is $O(\epsilon^3)$.

Adding I and II gives

(31)
$$\begin{aligned}
0 =& A(z, y)\partial_t R^a + (1 + \partial_t^2 y - \partial_t u K_0 \cdot)K_0\partial_\alpha R^u \\
&+ \partial_\alpha K_1(\Psi^z, \Psi^y)R^u + B
\end{aligned}$$

where

$$\begin{aligned}
B =& B_{II} + \partial_\alpha y[\partial_t^2, S_1(z, y)]R^u + B_{\mathrm{Res}} \\
\epsilon^{11/2} B_{\mathrm{Res}} =& (1 + Lz)\partial_t \epsilon^3 \Psi^a + \partial_\alpha y \partial_t^2(K(z, y)\Psi^u) \\
&+ (1 + \partial_t^2 y)\partial_\alpha \partial_t \Psi^y + \partial_t u \partial_\alpha \Psi^u.
\end{aligned}$$

The terms B_{II} and $\partial_\alpha y[\partial_t^2, S_1(z, y)]$ are small and smooth, and we can bound B_{Res} via the residual estimates, much as we did for N_1^u above. That is, we have

$$\|B\|_{s-1} \le C\left(\epsilon^3 + \epsilon^3 \|\bar{R}\|_{\mathfrak{H}_e^s} + \epsilon^{11/2}\|\bar{R}\|_{\mathfrak{H}_e^s}^2\right)$$

under the same hypotheses as in the above Lemmas.

At this time, it is tempting to simply invert $A(z, y)$. Though we could do this, the inverse of this operator is not smoothing. In particular the presence of the term $\partial_\alpha y K_0$ in A will cause problems. We can eliminate K_0 to highest order by letting $H_0(z, y) = (1 + Lz - \partial_\alpha y K_0)$ act on (31). We have for the first term

$$\begin{aligned}
&H_0(z, y)A(z, y)\partial_t R^a \\
=&(1 + Lz)^2 \partial_t R^u + (1 + Lz)\partial_\alpha y K_0 \partial_t R^a - \partial_\alpha y K_0\left((1 + Lz)\partial_t R^a\right) \\
&- \partial_\alpha y K_0\left(\partial_\alpha y K_0 \partial_t R^a\right) + H_0(z, y)(\partial_\alpha y S_1(z, y)\partial_t R^a) \\
=&\left((1 + Lz)^2 + (\partial_\alpha y)^2 + H_1(z, y)\cdot\right)\partial_t R^a
\end{aligned}$$

where

$$\begin{aligned}
H_1(z, y)\cdot =& \partial_\alpha y\left([Lz, K_0]\cdot - K_0[\partial_\alpha y, K_0]\cdot - (1 + K_0^2)\partial_\alpha y\cdot\right) \\
&+ H_0(z, y)\left(\partial_\alpha y S_1(z, y)\cdot\right).
\end{aligned}$$

Notice that H_1 is made up of smoothing operators, and is thus a smoothing operator.

Now, for the second term in (31) we have

$$H_0(z,y)(1 + \partial_t^2 y - \partial_t u K_0 \cdot) K_0 \partial_\alpha R^u$$

$$=(1 + Lz)(1 + \partial_t^2 y) K_0 \partial_\alpha R^u - (1 + Lz) \partial_t u K_0^2 \partial_\alpha R^u$$

$$- \partial_\alpha y K_0 \left((1 + \partial_t^2 y) K_0 \partial_\alpha R^u\right) + \partial_\alpha y K_0 \left(\partial_t u K_0^2 \partial_\alpha R^u\right)$$

$$=(1 + Lz)(1 + \partial_t^2 y) K_0 \partial_\alpha R^u$$

$$- \partial_\alpha y [K_0, \partial_t^2 y] K_0 \partial_\alpha R^u + \partial_\alpha y K_0 \left(\partial_t u K_0^2 \partial_\alpha R^u\right)$$

$$+ \left(\partial_t u(1 + Lz) + \partial_\alpha y(1 + \partial_t^2 y)\right) K_0^2 \partial_\alpha R^u.$$

Notice that by comparing the last line of the above with (29), we see that it is identically zero! One more rearrangement of this yields

$$\left((1 + Lz)(1 + \partial_t^2 y) - \partial_\alpha y \partial_t u\right) K_0 \partial_\alpha R^u + B_2$$

where

$$B_2 = \partial_\alpha y \left(-[K_0, \partial_t^2 y] + [K_0, \partial_t u] K_0 + \partial_t u(1 + K_0^2)\right) K_0 \partial_\alpha u$$

is a smooth and small function by Lemma 4 in Section 5.

If we let

$$f = ((1 + Lz)^2 + (\partial_\alpha y)^2)^{-1}$$

$$g = (1 + Lz)(1 + \partial_t^2 y) - \partial_\alpha y \partial_t u$$

then we have transformed (31) into

$$0 = (1 + f H_1(z,y)) \partial_t R^a + f g K_0 \partial_\alpha R^u$$

$$+ f H_0(z,y) \left(\partial_\alpha \left(K_1(\Psi^z, \Psi^y) R^u\right)\right)$$

$$+ f \left(B_2 + H_0(z,y) B\right).$$

By the Neumann series,

$$(1 + f H_1(z,y) \cdot)^{-1} = 1 + \sum_{n=0}^{\infty} (-1)^n f^n H_1^n(z,y) \cdot .$$

Since H_1 is smoothing, this is the identity plus a smoothing piece. Let

$$H_2(z,y) \cdot = \sum_{n=0}^{\infty} (-1)^n f^n H_1^n(z,y) \cdot .$$

Thus,

$$0 = \partial_t R^a + f g K_0 \partial_\alpha R^u$$

$$+ f H_0(z,y) \left(\partial_\alpha \left(K_1(\Psi^z, \Psi^y) R^u\right)\right) - N_1^a - N_2^a$$

where

$$-N_1^a = H_2(z,y) \left(f g K_0 \partial_\alpha R^u + f H_0(z,y) \left(\partial_\alpha \left(K_1(\Psi^z, \Psi^y) R^u\right)\right)\right)$$

$$-N_2^a = (1 + H_2(z,y)) f \left(B_2 + H_0(z,y) B\right).$$

Finally we rewrite the above as

$$\partial_t R^a = -\left(1 + K_0 a - Lz + N_s^a\right) \partial_\alpha K_0 R^u - \partial_\alpha \left(K_1(\Psi^z, \Psi^y) R^u\right) + N^a$$

with

$$N_s^a = fg - (1 + K_0 a - Lz)$$
$$N^a = N_1^a + N_2^a + N_3^a$$
$$N_4^a = - (f H_0(z,y) - 1) \left(\partial_\alpha \left(K_1(\Psi^z, \Psi^y) R^u \right) \right).$$

Notice that $(1 + K_0 a - Lz)$ is the first order approximation to fg. And thus, using techniques exactly like those we used in proving the bounds on N^y and N^u, we have the following.

Lemma 13. *For all $C_R > 0$, there exists ϵ_0 such that for all $\epsilon \in (0, \epsilon_0)$ and t such that $\sup_{0 \leq t' \leq t} \|\bar{R}_e(\cdot, t')\|_{\mathfrak{H}_e^s} \leq C_R$ we have*

$$\max \{ \|N_s^a\|_{s-1}, \|N^a\|_{s-1} \} \leq C \left(\epsilon^3 + \epsilon^3 \|\bar{R}_e\|_{\mathfrak{H}_e^s} + \epsilon^{11/2} \|\bar{R}_e\|_{\mathfrak{H}_e^s}^2 \right).$$

Recapping, we have shown that the three dimensional system may be rewritten as

$$\partial_t R^z = K_0 R^u$$

(32) $$\partial_t R^y = K_0 R^u + M_1(\Psi^z) \partial_\alpha R^u - (\Psi^y + K_0(\epsilon^2 \Psi^y K_0)) \partial_\alpha R^u + N^y$$

$$\partial_t R^u = -(1 - \epsilon^2 W_1) \partial_\alpha R^y + N^u$$

and the four dimensional system as

$$\partial_t R^z = K_0 R^u$$
$$\partial_t R^y = K_0 R^u + M_1(\Psi^z) \partial_\alpha R^u$$
$$\qquad - (\Psi^y + K_0(\epsilon^2 \Psi^y K_0)) \partial_\alpha R^u + N^y$$

(33)

$$\partial_t R^u = R^a$$
$$\partial_t R^a = - (1 + K_0 a - Lz + N_s^a) \partial_\alpha K_0 R^u$$
$$\qquad - \partial_\alpha \left(K_1(\Psi^z, \Psi^y) R^u \right) + N^a.$$

We remark now that these are only cosmetically different than the equations which determine the evolution of the error for the KdV approximation alone in [31]. See p 1524 for the equations in three dimensions and p 1526 in four. Their variables

$$(Z_1, X_2, U_1, V_1)$$

correspond to our

$$(z, y, u, a),$$

and their functions

$$(N^2, N^3, N^4, N^5, N^8)$$

are our

$$(N^z, N^y, N^u, N_s^a, N^a).$$

The only difference of note is that their estimates contain a term they call $q(t)$ while ours do not. This term, which is related to the interaction of the left and right moving wavetrains, has been removed in this paper by the inclusion of the function W_3 in the approximating functions Ψ. This simplification does not adversely affect the means which they employ to prove

that the error functions remain $O(1)$ over the long time scale. Therefore we appeal to their results on *pp* 1524-1533. That is,

Proposition 5. *For all $T_0 > 0$, $s > 4$ and $C_I > 0$, there exists ϵ_0 such that for all $0 \le \epsilon \le \epsilon_0$, the unique solution $\bar{R}_e$ of (33) with initial conditions such that*

$$\|\bar{R}_e(\cdot, 0)\|_{\mathfrak{H}_e^s} \le C_I$$

satisfies

$$\sup_{t \in [0, T_0 \epsilon^{-3}]} \|\bar{R}_e(\cdot, t)\|_{\mathfrak{H}_e^s} \le C$$

where C is independent of ϵ.

Implicit in the above Proposition is the assumption that the initial conditions for the water wave problem have the form:

$$\begin{pmatrix} z(\alpha, 0) \\ y(\alpha, 0) \\ u(\alpha, 0) \end{pmatrix} = \begin{pmatrix} \Psi^z(\alpha, 0) \\ \Psi^y(\alpha, 0) \\ \Psi^u(\alpha, 0) \end{pmatrix} + \epsilon^{11/2} \bar{R}_0(\alpha).$$

So we see that this Proposition immediately proves Theorem 3.

Now that we have this result, there are a few small steps, and one big step, needed to prove Theorem 2. The first simple step is to note that the z is not a very physical coordinate and that we would prefer estimates for x_α. Since L is a bounded operator and gives the relationship between both z and x_α and Ψ^z and Ψ^d, we have automatically

$$\sup_{t \in [0, T_0 \epsilon^{-3}]} \|x_\alpha(\cdot, t) - \Psi^d(\cdot, t)\|_{H^s} \le C\epsilon^{11/2}.$$

Secondly, the expressions for Ψ^d, Ψ^y and Ψ^u contain terms of $O(\epsilon^6)$. These terms were needed to make the residual sufficiently small, but they are unneccessary now. Moreover, the appearance of the operator L^{-1} and inverse derivatives in the definitions of Ψ^y and Ψ^u is not very intuitive. And so, it is a simple consequence of Lemma 1 and the triangle inequality that

$$\|\Psi^d - \psi^d\|_s \le C\epsilon^{11/2}$$
$$\|\Psi^y - \psi^y\|_s \le C\epsilon^{11/2}$$
$$\|\Psi^u - \psi^u\|_s \le C\epsilon^{11/2}$$

where ψ^d, ψ^y and ψ^u were given in the Introduction in equations (4), (5) and (6). And so we have the corollary

Corollary 1. *If the initial conditions for (WW) are of the form*

$$(34) \qquad \begin{pmatrix} x_\alpha(\alpha, 0) \\ y(\alpha, 0) \\ u(\alpha, 0) \end{pmatrix} = \begin{pmatrix} \psi^d(\alpha, 0) \\ \psi^y(\alpha, 0) \\ \psi^u(\alpha, 0) \end{pmatrix} + \epsilon^{11/2} \bar{R}_1(\alpha)$$

with $\|\bar{R}_1\|_{\mathfrak{H}^s} \leq C_I$ then the solution of (WW) satisfies the estimate:

$$\left\| \begin{pmatrix} x_\alpha(\cdot, t) \\ y(\cdot, t) \\ u(\cdot, t) \end{pmatrix} - \begin{pmatrix} \psi^d(\cdot, t) \\ \psi^y(\cdot, t) \\ \psi^u(\cdot, t) \end{pmatrix} \right\|_{\mathfrak{H}^s} \leq C_F \epsilon^{11/2}$$

for $t \in [0, T_0 \epsilon^{-3}]$. The constant C_F does not depend on ϵ.

Finally, we must deal with initial conditions. Recall from the discussion in Section 2 that it is typical to specify the initial data for the water wave problem in the long wavelength, small amplitude limit by

(35) $$(\bar{x}_{\bar{\alpha}}(\bar{\alpha}, 0), \bar{y}(\bar{\alpha}, 0), \bar{u}(\bar{\alpha}, 0)) = (0, \epsilon^2 \Theta_y(\epsilon\bar{\alpha}), \epsilon^2 \Theta_u(\epsilon\bar{\alpha})).$$

However, the above results are applicable if the initial data is of the form seen in (34). We eliminate this discrepency by altering the initial parameterization of the free surface. What should this change be? Clearly,

(36) $$\bar{\alpha} = \alpha + x(\alpha, 0).$$

Now set $U(\beta, 0) = U_0(\beta)$, $V(\beta, 0) = V_0(\beta)$, $F(\beta, 0) = F_0(\beta)$, $G(\beta, 0) = G_0(\beta)$, and $P(\beta, 0) = 0$, and let

$$\begin{aligned} \bar{\alpha} &= \alpha + \int_0^\alpha \psi^d(a, 0) da \\ &= \alpha + \epsilon X_1(\epsilon\alpha) + \epsilon^3 X_2(\epsilon\alpha) \end{aligned}$$

where

(37)
$$\begin{aligned} \epsilon X_1(\alpha) &= -\epsilon^2 \int_0^\alpha (U_0(\epsilon a) + V_0(\epsilon a)) \, da \\ \epsilon^3 X_2(\alpha) &= -\epsilon^4 \int_0^\alpha (F_0(\epsilon a) + G_0(\epsilon a)) \, da. \end{aligned}$$

With this definition, we clearly have satisfied the first condition in (34). We also want

$$\begin{aligned} \Theta_y(\epsilon\bar{\alpha}) &= \epsilon^2 \psi^y(\alpha) + O(\epsilon^{11/2}) \\ \Theta_u(\epsilon\bar{\alpha}) &= \epsilon^2 \psi^u(\alpha) + O(\epsilon^{11/2}) \end{aligned}$$

or rather

$$\begin{aligned} \epsilon^2 \Theta_y(\epsilon\alpha + \epsilon^2 X_1(\epsilon\alpha) + \epsilon^4 X_2(\epsilon\alpha)) &= \epsilon^2 \psi^y(\alpha) + O(\epsilon^{11/2}) \\ \epsilon^2 \Theta_u(\epsilon\alpha + \epsilon^2 X_1(\epsilon\alpha) + \epsilon^4 X_2(\epsilon\alpha)) &= \epsilon^2 \psi^u(\alpha) + O(\epsilon^{11/2}). \end{aligned}$$

Applying Taylor's theorem we have

$$\begin{aligned} \Theta_y + \epsilon^2 X_1 \Theta_y' &= (U_0 + V_0) + \epsilon^2 (\frac{1}{3}\partial_{\beta_-}^2 U_0 + \frac{1}{3}\partial_{\beta_+}^2 V_0) \\ &\quad + \epsilon^2 (F_0 + G_0) + \epsilon^2 (U_0 + V_0)^2 \\ \Theta_u + \epsilon^2 X_1 \Theta_u' &= (U_0 - V_0) + \epsilon^2 (\frac{1}{6}\partial_{\beta_-}^2 U_0 - \frac{1}{6}\partial_{\beta_+}^2 V_0) \\ &\quad + \epsilon^2 (F_0 - G_0) + \epsilon^2 (\frac{3}{4}U_0^2 - \frac{3}{4}V_0^2). \end{aligned}$$

We can solve the above by taking

(38)
$$U_0 = 1/2(\Theta_y + \Theta_u)$$
$$V_0 = 1/2(\Theta_y - \Theta_u)$$

and

(39)
$$F_0 = 1/2(h_y + h_u)$$
$$G_0 = 1/2(h_y - h_u)$$

where

$$h_y = X_1\Theta'_y - \frac{1}{3}\partial^2_{\beta_-}U_0 - \frac{1}{3}\partial^2_{\beta_+}V_0 - (U_0 + V_0)^2$$

$$h_u = X_1\Theta'_u - \frac{1}{6}\partial^2_{\beta_-}U_0 + \frac{1}{6}\partial^2_{\beta_+}V_0 - \frac{3}{4}U_0^2 + \frac{3}{4}V_0^2.$$

The functions U_0, V_0, F_0 and G_0 are all in $H^s(4)$, and so the use of Taylor's theorem is justified by Lemma 10. So we have proven:

Lemma 14. *Given initial conditions for the water wave equation in the form (35), define U_0, V_0, F_0, G_0, X_1 and X_2 as in (38), (39) and (37). Then the reparameterization of the initial profile given by:*

$$\bar{\alpha} = \alpha + \epsilon X_1(\epsilon\alpha) + \epsilon X_2(\epsilon\alpha)$$

results in initial conditions given by (34).

Remark 10. *Let $\varphi^\pm(\beta_\pm, 0) = \varphi_0^\pm(\beta_\pm)$. Then this Lemma will still be true if we replace F_0 with φ_0^- and G_0 with φ^+ and set F_0 and G_0 to be identically zero. That is, we have some choice in the way we select the initial conditions for the higher order equations.*

Combining this Lemma with Corollary 1 we prove Theorem 2. So we are done.

7. Assorted proofs

Proof. **For Lemma 8**: Let $f^{-1}(\xi) = \xi - g_2(\xi)$. Since $f^{-1}(f(\alpha)) = \alpha$ we have

$$\alpha = f(\alpha) - g_2(f(\alpha)),$$

or rather

(40)
$$g_2(f(\alpha)) = g(\alpha)$$

Notice that this relation implies $\|g_2\|_{L^\infty} = \|g\|_{L^\infty}$. Taking a derivative, we have

$$g'_2(f(\alpha)) = \frac{g'(\alpha)}{1 + g'(\alpha)}$$

which implies that $\|g'_2\|_{L^\infty} \leq C\|g'\|_{L^\infty}$. If we expand the left hand side of (40) by the mean value theorem we see

$$g_2(\alpha + g(\alpha)) = g(\alpha)$$
$$g_2(\alpha) + g'_2(\alpha^*)g(\alpha) = g(\alpha).$$

This implies $g_2(\alpha) = g(\alpha) + O(\|g'\|_{L^\infty}\|g\|_{L^\infty})$. Now, (40) can be rewritten and expanded using Taylor's theorem:

$$\begin{aligned}
g_2(\xi) &= g(f^{-1}(\xi)) \\
&= g(\xi - g(\xi) + O(\|g'\|_{L^\infty}\|g\|_{L^\infty})) \\
&= g(\xi) + g'(\xi)(-g(\xi) + O(\|g'\|_{L^\infty}\|g\|_{L^\infty})) \\
&\quad + 1/2 g''(\xi^*)(-g(\xi) + O(\|g'\|_{L^\infty}\|g\|_{L^\infty}))^2 \\
&= g(\xi) - g(\xi)g'(\xi) - E
\end{aligned}$$

which completes the proof. $\square$

Proof. **For Lemma 9**: Since $(1 + \beta^2)^{n/2}F(\beta) \in H^s$, by the Sobolev embedding Theorem there is a C such that

$$F(\beta) \le C(1 + \beta^2)^{-n/2}.$$

So

$$\begin{aligned}
&\int_{|\alpha|>C_0\epsilon^{-3}} |F(\epsilon\alpha)|^2 d\alpha \\
&\le C \int_{|\alpha|>C_0\epsilon^{-3}} |1 + (\epsilon\alpha)^2|^{-n} d\alpha \\
&\le C \int_{|\alpha|>C_0\epsilon^{-3}} |\epsilon\alpha|^{-2n} d\alpha \\
&\le C\epsilon^{-2n} \int_{|\alpha|>C_0\epsilon^{-3}} |\alpha|^{-2n} d\alpha \\
&\le C\epsilon^{-2n}(\epsilon^{-3})^{-2n+1} \\
&\le C\epsilon^{4n-3}.
\end{aligned}$$

The higher derivatives are bounded in exactly the same fashion. The extra powers of ϵ come from the long wavelength scaling. $\square$

Proof. **For Lemma 1**: The proof is a straightforward calculation.

$$\begin{aligned}
&\|Af(\cdot) - A_n f(\cdot)\|_s^2 \\
&= \int (1+k^2)^s |(\widehat{A}(k) - \widehat{A_n}(k))\widehat{f}(k)|^2 dk \\
&\le C \int (1+k^2)^s |k^n \widehat{f}(k)|^2 dk \\
&= C \int (1+k^2)^s |\widehat{\partial_x^n f}(k)|^2 dk \\
&= C\|\partial_x^n f(\cdot)\|_s^2.
\end{aligned}$$

The proof for long wavelength data follows immediately from this. $\square$

Proof. **For Lemma 2**: The fact that $g(\epsilon x)$ is bounded as such in L^2 follows automatically from the geometric series approximation. That is, since $|\epsilon^2 f(\epsilon x)| \leq 1/2$, we know that:

$$|g(\epsilon x)| \leq C|\epsilon^2 f(\epsilon x)|^2.$$

And thus we have:

$$\|g(\epsilon \cdot)\|^2_{L^2}$$
$$\leq C \int \epsilon^4 |f(\epsilon x)|^4 dx$$
$$\leq C\epsilon^4 \|f(\cdot)\|^2_{L^\infty} \int f^2(\epsilon x) dx$$
$$\leq C\epsilon^{7/2} \|f(\cdot)\|_s.$$

Now consider the L^2 norm of $g'(\epsilon x)$. A direct calculation shows that

$$\frac{d}{dx} g(\epsilon x) = -\epsilon^3 f'(\epsilon x) \left(\frac{1}{(1 + \epsilon^2 f(\epsilon x))^2} - 1 \right).$$

Taylor's theorem shows that

$$|\frac{d}{dx} g(\epsilon x)| \leq C|\epsilon^3 f'(\epsilon x)||\epsilon^2 f(\epsilon x)|.$$

And so, just as before we have that this is bounded by $C\epsilon^{9/2}$ (which is of course bounded by $C\epsilon^{7/2}$).

We could keep on going in this fashion—showing each derivative of g is bounded. This is however difficult as finding higher and higher derivatives is a notationally taxing job—see the expression of Faa-di-Bruno for proof of that!

Instead we take the following approach. Let

$$h(y) = \frac{1}{1+y} - 1 + y.$$

For $y \in [-1/2, 1/2]$, this function is real analytic and there exists another function $\tilde{h}(y)$ (real and analytic on the same interval) such that $h(y) = y\tilde{h}(y)$. Now, define $\tilde{h}_\epsilon(Y) = \tilde{h}(\epsilon^2 y)$. We have that:

$$\|\tilde{h}_\epsilon(\cdot)\|_{C^s[-1/2\epsilon^{-2}, 1/2\epsilon^{-2}]} \leq \|\tilde{h}(\cdot)\|_{C^s[-1/2, 1/2]}.$$

The point here is that the C^s norm of $\tilde{h}_\epsilon$ can be bounded *independently* of ϵ.

Now notice that $g(X) = \epsilon^2 f(X)\tilde{h}_\epsilon(f(X))$. Since $f \in H^{s+1}$, we know that $f \in C^s$. This implies that $\tilde{h}_\epsilon(f(X)) \in C^s$, with C^s norm bounded independent of ϵ. Thus we have $f(X)\tilde{h}_\epsilon(f(X)) \in H^s$, with a norm bounded independent of ϵ. With this in hand, we have that $\|g(\cdot)\|_s \leq C\epsilon^2$, with

$C \neq C(\epsilon)$. Now, the derivatives of g can be bounded as follows:

$$\|\frac{d^n}{dx^n}g(\epsilon\cdot)\|_{L^2}$$
$$=\epsilon^n\|g^{(n)}(\epsilon\cdot)\|_{L^2}$$
$$\leq C\epsilon^{n-1/2}\|g(\cdot)\|_s$$
$$\leq C\epsilon^{n+3/2}.$$

Provided $n \geq 2$, this term is small enough. And so we have shown that the first s derivatives are sufficiently small in L^2, and we have proved the Lemma. $\qquad\square$

References

[1] Walid Ben Youssef and Thierry Colin. Rigorous derivation of Korteweg-de Vries-type systems from a general class of nonlinear hyperbolic systems. *M2AN Math. Model. Numer. Anal.*, 34(4):873–911, 2000.

[2] J. L. Bona, W. G. Pritchard, and L. R. Scott. An evaluation of a model equation for water waves. *Philos. Trans. Roy. Soc. London Ser. A*, 302(1471):457–510, 1981.

[3] Jerry L. Bona and Min Chen. A Boussinesq system for two-way propagation of nonlinear dispersive waves. *Phys. D*, 116(1-2):191–224, 1998.

[4] Jerry L. Bona, David Lannes, and Thierry Colin. Long-waves appoximations for water-waves. Preprint, 2003.

[5] J. G. B. Byatt-Smith. An integral equation for unsteady surface waves and a comment on the Boussinesq equation. *J. Fluid Mech.*, 49:625–633, 1971.

[6] J. G. B. Byatt-Smith. The reflection of a solitary wave by a vertical wall. *J. Fluid Mech.*, 197:503–521, 1988.

[7] J. G. B. Byatt-Smith. The head-on interaction of two solitary waves of unequal amplitude. *J. Fluid Mech.*, 205:573–579, 1989.

[8] M. J. Cooker, P. D. Weidman, and D. S. Bale. Reflection of a high-amplitude solitary wave at a vertical wall. *J. Fluid Mech.*, 342:141–158, 1997.

[9] Walter Craig. An existence theory for water waves and the Boussinesq and Korteweg-de Vries scaling limits. *Comm. Partial Differential Equations*, 10(8):787–1003, 1985.

[10] Walter Craig and Mark D. Groves. Hamiltonian long-wave approximations to the water-wave problem. *Wave Motion*, 19(4):367–389, 1994.

[11] P. G. Drazin and R. S. Johnson. *Solitons: an introduction.* Cambridge University Press, Cambridge, 1989.

[12] Holger R. Dullin, Georg A. Gottwald, and Darryl D. Holm. Camassa-Holm, Korteweg-de Vries-5 and other asymptotically equivalent equations for shallow water waves. *Fluid Dynam. Res.*, 33(1-2):73–95, 2003. In memoriam Prof. Philip Gerald Drazin 1934–2002.

[13] J. D. Fenton and M. M. Rienecker. A Fourier method for solving nonlinear water-wave problems: application to solitary-wave interactions. *J. Fluid Mech.*, 118:411–443, 1982.

[14] Mariana Haragus, David P. Nicholls, and David H. Sattinger. Solitary wave interactions of the Euler-Poisson equations. *J. Math. Fluid Mech.*, 5(1):92–118, 2003.

[15] M. Hărăguş-Courcelle and D. H. Sattinger. Inversion of the linearized Korteweg-de Vries equation at the multi-soliton solutions. *Z. Angew. Math. Phys.*, 49(3):436–469, 1998.

[16] L. A. Kalyakin. Long-wave asymptotics. Integrable equations as the asymptotic limit of nonlinear systems. *Uspekhi Mat. Nauk*, 44(1(265)):5–34, 247, 1989.

[17] Tadayoshi Kano and Takaaki Nishida. Sur les ondes de surface de l'eau avec une justification mathématique des équations des ondes en eau peu profonde. *J. Math. Kyoto Univ.*, 19(2):335–370, 1979.

[18] Tadayoshi Kano and Takaaki Nishida. A mathematical justification for Korteweg-de Vries equation and Boussinesq equation of water surface waves. *Osaka J. Math.*, 23(2):389–413, 1986.

[19] Pius Kirrmann, Guido Schneider, and Alexander Mielke. The validity of modulation equations for extended systems with cubic nonlinearities. *Proc. Roy. Soc. Edinburgh Sect. A*, 122(1-2):85–91, 1992.

[20] Yuji Kodama. On integrable systems with higher order corrections. *Phys. Lett. A*, 107(6):245–249, 1985.

[21] Yuji Kodama and Tosiya Taniuti. Higher order corrections to the soliton-velocity and the linear dispersion relation. *Phys. Scripta*, 20(3-4):486–489, 1979.

[22] David Lannes. Secular growth estimates for hyperbolic systems. *J. Differential Equations*, 190(2):466–503, 2003.

[23] David Lannes. Well-posedness of the water-waves equations. Preprint, 2003.

[24] T. Maxworthy. Experiments on collisions between solitary waves. *J. Fluid Mech.*, 76:177–185, 1976.

[25] J. W. Miles. Obliquely interacting solitary waves. *J. Fluid Mech.*, 79:157–169, 1977.

[26] Peter J. Olver. Hamiltonian and non-Hamiltonian models for water waves. In *Trends and applications of pure mathematics to mechanics (Palaiseau, 1983)*, volume 195 of *Lecture Notes in Phys.*, pages 273–290. Springer, Berlin, 1984.

[27] Robert L. Sachs. Completeness of derivatives of squared Schrödinger eigenfunctions and explicit solutions of the linearized KdV equation. *SIAM J. Math. Anal.*, 14(4):674–683, 1983.

[28] Robert L. Sachs. A justification of the KdV approximation to first order in the case of N-soliton water wavers in a canal. *SIAM J. Math. Anal.*, 15(4):468–489, 1984.

[29] D. H. Sattinger and Yi Li. Matlab codes for nonlinear dispersive wave equations. 1998.

[30] Guido Schneider. The long wave limit for a Boussinesq equation. *SIAM J. Appl. Math.*, 58(4):1237–1245 (electronic), 1998.

[31] Guido Schneider and C. Eugene Wayne. The long-wave limit for the water wave problem. I. The case of zero surface tension. *Comm. Pure Appl. Math.*, 53(12):1475–1535, 2000.

[32] Guido Schneider and C. Eugene Wayne. Kawahara dynamics in dispersive media. *Phys. D*, 152/153:384–394, 2001. Advances in nonlinear mathematics and science.

[33] Guido Schneider and C. Eugene Wayne. The rigorous approximation of long-wavelength capillary-gravity waves. *Arch. Rational Mech. Anal.*, 162(3):247–285, 2002.

[34] C. H. Su and Rida M. Mirie. On head-on collisions between solitary waves. *J. Fluid Mech.*, 98(3):509–525, 1980.

[35] C. H. Su and Rida M. Mirie. Collisions between two solitary waves. part 2. a numerical study. *J. Fluid Mech.*, 115:475–492, 1982.

[36] C. Eugene Wayne and J. Douglas Wright. Higher order modulation equations for a Boussinesq equation. *SIAM J. Appl. Dyn. Sys.*, 1:271–302, 2002.

[37] J. Douglas Wright. Higher order corrections to the KdV approximation for water waves. 2003.

[38] Sijue Wu. Well-posedness in Sobolev spaces of the full water wave problem in 2-D. *Invent. Math.*, 130(1):39–72, 1997.

[39] Sijue Wu. Well-posedness in Sobolev spaces of the full water wave problem in 3-D. *J. Amer. Math. Soc.*, 12(2):445–495, 1999.

[40] Qisu Zho and Chau-Hsing Su. Overtaking collision between two solitary waves. *Phys. Fluids*, 29:2113–2123, 1986.